FROM: AFIT/CI

21 July 1989

SUBJECT: Review of Thesis/Disseration for Public Release

TO: PA

1. Request you review the attached for public release prior to being sent to DTIC.

2. Reply by indorsement to CI NLT ______________.

Ernest A. Haygood

ERNEST A. HAYGOOD, 1st Lt, USAF
Executive Officer
Civilian Institution Programs

1 Atch
DISSERTATION 89-079
SWEHOSKY

1st Ind, AFIT/PA

08 FEB 1990

TO: CI

Approved/~~Disapproved~~ for public release.
Log Number: 89-10-123

Harriet D Moultrie

HARRIET D. MOULTRIE, Capt, USAF
Director, Office of Public Affairs

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS NONE
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFIT/CI/CIA-89-079
6a. NAME OF PERFORMING ORGANIZATION AFIT STUDENT AT UNIV OF TX-Austin 6b. OFFICE SYMBOL *(If applicable)*	7a. NAME OF MONITORING ORGANIZATION AFIT/CIA
6c. ADDRESS *(City, State, and ZIP Code)*	7b. ADDRESS *(City, State, and ZIP Code)* Wright-Patterson AFB OH 45433-6583
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL *(If applicable)*	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS *(City, State, and ZIP Code)*	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE *(Include Security Classification)* (UNCLASSIFIED)
A COGNITIVE MODEL OF COLLEGE MATHEMATICS PLACEMENT

12. PERSONAL AUTHOR(S)
FRANK JOSEPH SWEHOSKY

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT *(Year, Month, Day)*	15. PAGE COUNT
THESIS/DISSERTATION	FROM ____ TO ____	1989	175

16. SUPPLEMENTARY NOTATION
APPROVED FOR PUBLIC RELEASE IAW AFR 190-1
ERNEST A. HAYGOOD, 1st Lt, USAF
Executive Officer, Civilian Institution Programs

17. COSATI CODES			18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL ERNEST A. HAYGOOD, 1st Lt, USAF	22b. TELEPHONE *(Include Area Code)* (513) 255-2259 22c. OFFICE SYMBOL AFIT/CI

DD Form 1473, JUN 86 *Previous editions are obsolete.* SECURITY CLASSIFICATION OF THIS PAGE

AFIT/CI "OVERPRINT"

A COGNITIVE MODEL OF COLLEGE MATHEMATICS PLACEMENT

APPROVED BY
SUPERVISORY COMMITTEE:

L Ray Carry

[illegible]

Ralph W. Cain

Thomas F. Curry

H. Paul Kelley

Soli Deo Gloria

A COGNITIVE MODEL OF COLLEGE MATHEMATICS PLACEMENT

by

FRANK JOSEPH SWEHOSKY, B.S., M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

AUGUST, 1989

ACKNOWLEDGMENTS

I was blessed with the support of several people, organizations, and institutions; too many to individually recognize. I wish however, to identify those who especially aided me with this endeavor.

Five individuals, my dissertation committee, were most helpful. My deepest gratitude goes to L. Ray Carry, the chairman of this committee. He accepted the burden of my professional development, not just the publication of this document. Under his tutelage I matured as a mathematics educator, a behavioral scientist, an academician, and a Christian. Also, the guidance that H. Paul Kelley provided, based on his experience with placement and standard setting, was invaluable.

The other members of my committee also made important contributions to this project: Peter W. M. John performed the vital role of monitoring critical elements of quality control, Ralph W. Cain's attention to detail helped ensure the readability of this document, and Lt Colonel Thomas F. Curry acted as my liaison (and primary benefactor) with the United States Air Force Academy. The synergism created from the efforts of this team contributed to the final product; which after all, was not this dissertation.

Special thanks are extended to the Chairman and other officers, past and present, in the Department of Mathematical Sciences at the United States Air Force Academy. This project was made possible only by their steadfast institutional support and individual cooperation.

The most important factor to my success however, was my family. My wife's unceasing love, emotional balance, and empathy was the foundation from which I was able to begin and persevere to

the completion of this program. I must happily report that it will take more than a lifetime to repay this debt to her. My children also provided unfailing love, vitality, and enthusiasm which continually reminded me of what was most important.

A COGNITIVE MODEL OF COLLEGE MATHEMATICS PLACEMENT

Publication No. __________

Frank Joseph Swehosky, Ph.D.
The University of Texas at Austin, 1989

Supervising Professor: L. Ray Carry

This study developed and validated the Cognitive Model of College Mathematics Placement and compared its effectiveness to that of Willingham's (1974) vertical placement model as well as the two empirical models. The Cognitive model was based on Skemp's (1979) theory of intelligent learning and Wilson's (1971) Model of Mathematics Achievement. The Cognitive model was validated using historical mathematics placement data from the graduating classes of 1989, 1990, and 1991 at the United States Air Force Academy. The study focused on the precalculus -- calculus placement decision.

The Cognitive model uses novel, or analysis level, placement test items in an attempt to assess the degree of connectedness of students' schemas and non-analysis items to assess the degree of accuracy and completeness of their schemas relative to the requirements of a precalculus course. Placement test scores may be partitioned to give analysis and non-analysis subtest scores which can then be used to predict students' achievement in calculus. Students are placed into precalculus if their predicted final calculus grades are below the cutoff score identified with the methods of Appenzellar and Kelley (1983).

Some of the conclusions of this study were:

1. A Cognitive Model of College Mathematics Placement could be developed from Skemp's theory of intelligent learning and Wilson's Model of Mathematical Achievement with locally developed placement examinations providing the predictor variables.

2. For the Classes of 1989, 1990, and 1991 at the U.S. Air Force Academy, the Cognitive model was a marginally valid placement system. The cognitive subscales were content valid but not, in general, reliable. In addition, confirmatory factor analysis did not provide any significant empirical support for the cognitive classification of the placement test items derived from expert opinion. Reasonable levels of predictive validity of the cognitive model were observed for the Classes of 1990 and 1991; however, different sets of cognitive variables were significant predictors of final calculus grades. The number of significant cognitive predictors increased with the number of unsuccessful students in calculus so that all the cognitive variables were significant predictors of final calculus grades for the Class of 1991.

3. In practical terms, the various placement models displayed the same levels of effectiveness. The Willingham and the empirical placement models consistently produced a small increase in the number of correct placements, but the Cognitive model consistently provided a significantly better prediction of final calculus grades.

TABLE OF CONTENTS

Page

Page

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Statement of Problem

Many decisions face a college or university with each entering freshmen class, and the United States service academies are no different. One of the earliest decisions made at the United States Air Force Academy (USAFA) is the course placement of freshmen within the core mathematics sequence. Presently the Department of Mathematical Sciences (DFMS) at the USAFA uses a variety of data and techniques to perform this placement. No learning theoretic rationale has been articulated to support the use of the current techniques.

Learning theoretic rationales for mathematics placement procedures are rare in the research literature despite the abundance of studies about college mathematics placement. Willingham's (1974) model of placement is one of the most frequently cited learning theoretic rationales.

Willingham (1974) developed 12 models of assignment, selection, placement, and exemption. These models were based on decision theory and Gagné's (1970) theory of instruction. Willingham's most applicable model for placement in a college precalculus-differential calculus sequence is his vertical placement model. The basic elements of this model applied in this context include (a) constructing a placement examination directly tied to the precalculus course objectives to determine the students' levels of achievement within the precalculus-calculus sequence; (b) randomly placing students into the sequences; and (c) using trait-treatment interaction (TTI) techniques (Cronbach & Gleser, 1957) to establish the validity of the placement test using the final grades of the common course in the sequence as the outcome variable. When TTI techniques cannot be used, Willingham suggested using other methods to establish the validity of the test as a placement tool.

One of the shortcomings of Willingham's placement models is that placement examinations' correlations with final course grades commonly range between .40 and .60 (Aleamoni, 1979). Thus, while a placement examination may yield different slopes of the regression lines for students enrolled in the two sequences, i.e., differential placement, the predictive power of the test may be quite low.

Although investigators have studied the effectiveness of a variety of prediction models (Gamache & Novick, 1985; Ervin, Hogrebe, Dwinell, & Newman, 1984; Bingham, 1972; Dunn, 1966), few have reported the theoretic justification for the connection between the predictors used and final course grades. Instead, most of these prediction models were selected empirically on the basis of predictive power.

Conspicuously absent from the rationale of placement systems discussed in the literature are the recent advances in the area of psychology termed "cognitive science" (Schoenfeld,1987). The emerging information processing and schemata theories may explain much of the student behavior that is observed in school mathematics. Of particular interest are the theories of Skemp (1979, 1987) concerning mathematical learning. Skemp's theories suggest that while prerequisite knowledge is critical for forming new concepts, it is also important that the prerequisite knowledge be adaptable to new situations (Skemp, 1987). For knowledge to be adaptable, the concepts must have been learned in a manner that facilitates assimilation of new concepts, reconstruction of existing concepts, and generalization of concepts to similar but unfamiliar situations.

Thus, what is needed to predict success in learning mathematics is not only a measure of the completeness of the prerequisite knowledge, but also a measure of the adaptability of the knowledge. Skemp (1987) further suggested that the best way to measure the adaptability of students' knowledge is through conducting diagnostic interviews or through presenting the students with situations that require application of knowledge in novel settings. Since diagnostic interviews are usually one-to-one interviews, this technique is impractical for the placement of large entering freshmen classes. Alternatively, Skemp argued that appropriate novel items can provide the needed assessments. Wilson's (1971) model of mathematics achievement provides a theoretical basis and suggests techniques for measuring knowledge in novel settings.

Research Purpose and Questions

Two objectives of this study were (a) to establish and validate a cognitive science model for college mathematics placement; and (b) to improve the mathematics placement procedures for the precalculus-calculus sequence at the USAFA. To this end, the following research questions were investigated:

I. Can a cognitive model for college mathematics placement be developed from Skemp's theory of learning and Wilson's model of mathematics achievement?

II. Which predictor variables are logically consistent with such a model for college mathematics placement?

III. Does the Cognitive Model for College Mathematics Placement, using the predictor variables identified in II, produce a valid placement procedure?

IV. Does the Cognitive Model for College Mathematics Placement produce a more effective placement than either the Willingham model or the two empirically-based placement models used by the US Air Force Academy?

Definitions

The following operational definitions were used by the investigator in conducting this research:

American College Testing Program Assessment (ACT): A basic intellectual skills test to predict success in college. The testing program contains four subtests: English Usage (ACT-E), Mathematics Usage (ACT-M), Social Studies Reading, (ACT-S), and Natural Sciences Reading (ACT-N). This test is frequently used by universities as an admission and placement tool.

Correct placement: The agreement between the observed and predicted success of students; a hit. For example, a student is said to be correctly placed if predicted and observed to be successful, or predicted and observed to be unsuccessful. The student is incorrectly placed if predicted to be successful and observed to be unsuccessful, or predicted to be unsuccessful and observed to be successful.

Cutoff score: The score on a mathematics test chosen to predict successful and unsuccessful categories of students.

Course sequence: An academic treatment in which college mathematics courses must be taken in a particular order. In this research there was a prerequisite course (precalculus), which covered prerequisite knowledge necessary for success in the subsequent criterion course (calculus). The two-course sequence was the precalculus course followed by the differential calculus course. This was referred to as the long sequence. Course sequences may also consist of a solitary course. Thus, some students were placed directly into the differential calculus course, i.e., the short sequence.

Efficient placement: The result of applying placement procedures such that the TTI observed using one placement procedure was significantly larger than that found by using another placement procedure. When a TTI is not observed because of limitations, an alternative definition may be used; the result when the number of students correctly placed under one placement procedure was larger than that observed from using the other procedure. Also, the results of a generalized F-test may support the evidence of an efficient placement.

Mathematics achievement: A multivariate psychological construct described by a content-by-cognitive behavior matrix. The levels of cognitive behavior are computation, comprehension, application, and analysis as defined by Wilson (1971). Analysis items are items that require students to apply knowledge, procedures, and algorithms in a novel problem setting. Non-analysis items are items that are not analysis items.

Novel problem: A type of test item whose solution has neither been taught nor practiced. Novel problems are not defined by their difficulty; rather they are defined in terms of the cognitive requirements necessary for their correct solution.

Placement: The enrollment of students into course sequences of different lengths that have a common outcome criterion, end-of-course-sequence achievement, in order to maximize the likelihood of success of the students. Vertical placement is the type of placement used when students are permitted to place out of one or more of the prerequisite courses in a fixed sequence of courses.

The College Board Scholastic Aptitude Test (SAT): A test of general intellectual skills and knowledge that are perceived to have been developed both in and out of high school. The test is composed of a Verbal (SAT-V) section and a Mathematical (SAT-M) section. This test, as the ACT, is commonly used as a tool for admission and placement at universities.

Success in a course sequence: The result when a student receives a final grade of at least a C- in the criterion course.

Trait-treatment interaction (TTI): The statistical result found when a prediction model, using a trait measure to predict achievement, is significantly improved by including a trait-by-treatment vector.

Validity: The ability of a test to measure the trait (or combination of traits) that it is designed to measure. Several types of validity are usually investigated to establish the usefulness of a particular test or set of procedures for placement purposes. Content validity determines if a test contains items that adequately measure the appropriate performance domain. Criterion-related validity provides information about the relationship between test scores and criterion scores of the trait; one type of criterion-related validity is predictive validity, the ability of test scores to predict criterion scores (e.g., final course grades). Another type of validity is construct validity, which supports using a given test as an appropriate measure of a psychological construct.

CHAPTER II

REVIEW OF RELATED RESEARCH LITERATURE

Introduction

Research Questions I and II call for constructing a new placement rationale and a model for implementing placement procedures in a college mathematics sequence based on cognitive science learning theory. In support of these two questions, four major areas of related research literature were reviewed in this section: general placement theory, cognitive science learning theory, measurement theory as applied to mathematics, and previous investigations in the area of mathematics placement. Also included in this section is a description of the proposed cognitive model for placement in college mathematics.

The review of general placement theory emphasizes the work of Cronbach and Gleser (1957) and Willingham (1974). These two works establish the placement problem and provide the fundamental rationale and techniques for developing and validating placement instruments and procedures.

Willingham (1974) based his rationale for placement on Gagné's (1970) learning theory. An alternate view of learning from cognitive science is reviewed and argued to be useful in developing a new rationale for mathematics placement. Three cognitive science theories are briefly described: constructivism, information processing, and schema theory. Skemp's (1979) theory of learning incorporates these basic theories and seems, to this investigator, to be particularly applicable to mathematics placement.

Skemp's (1979) theory provides a basis for describing successful mathematical behavior but does not provide operational definitions to construct and validate a placement examination. Wilson's (1971) model of mathematics achievement provides the technology for measuring the types of behavior specified by Skemp. These two theories are synthesized to form the proposed Cognitive Model of College Mathematics Placement. This synthesis includes a discussion of the theoretical basis and operational procedures of the model in the context of a precalculus-differential calculus two-course sequence.

Next, several dependent and independent variables, commonly used in previous college mathematics placement and prediction studies, are evaluated with respect to their consistency with the Cognitive model. This evaluation determined a set of independent and dependent variables that may be used in placing students into a college precalculus-calculus sequence.

General Placement Theory

Thorndike (1949) wrote one of the earliest books concerning placement in which he methodically discussed the military personnel placement issues of World War II. His methods continued to be developed by persons in industry and business. Cronbach and Gleser (1957, 1965) produced one of the first works that seriously analyzed placement within the education setting. Willingham described their work as a "classic and most comprehensive treatment of the logic and psychometric characteristics of the placement problem" (Willingham, 1973, p. 99) to that date.

Cronbach and Gleser (1957) defined placement as the process by which persons within an institution are put into different treatments. They rigorously formulated the mathematics procedures for placement using classical decision theory in a variety of settings. A key assumption in their procedure was that the treatments are dependent upon a single aptitude factor. This assumption permits aspects of the decision problem to be separated according to the nature of the test and the differences between competing treatments. Specifically, Cronbach and Gleser assumed that "only a single aptitude dimension is required to account for all communality between test scores and payoffs" (Cronbach and Gleser, p. 38).

A typical placement setting for college mathematics is where there are fixed course sequences with variable enrollments within the sequences. For this setting Cronbach and Gleser suggested that the ideal placement strategy is to maximize the payoff, usually a grade, for each individual student. The cutoff for determining placement is found through the technique of trait-treatment interaction (TTI). This technique requires the grade function for competing treatments to be graphed on a grade-by-trait surface. The optimum a priori strategy for placement is to select the point that maximizes the grade function.

Cronbach and Gleser proposed two criteria for judging the utility of a placement test, "The power of the test to measure the aptitude dimension s, and the power of s to predict differential payoff [differential prediction]" (Cronbach & Gleser, 1957, p. 68). The TTI techniques would then be used to establish the differential prediction (payoff) of the test as well as determine the cutoff for future placements.

Cronbach and Gleser suggested caution when using measures of general ability in placement systems, a warning that has been repeated through the literature.

> Tests presently used may be ineffective for placement even though they are good predictors within a treatment. Possibly quite different types of items would make superior placement tests, because qualities which determine differential response to various treatments are not generally those which best predict criterion performance within one treatment. General

> mental ability, for example, is likely to be correlated with success in mathematics no matter how the subject is taught. . . . A measure which predicted success under one treatment and not the other would be a much better aid to placement than a measure which predicts both (Cronbach & Gleser, 1957, p. 68).

Later Cronbach (1971) defended the basic TTI approach against a general regression approach when judging the usefulness of a placement test.

> The utility of a test for placement depends on the difference in regression slopes - not on the slopes or the correlation directly. A 'validity coefficient' indicating that test X predicts success within a treatment <u>tells nothing about its usefulness for placement</u> (Cronbach, 1971, p.500).

It follows that a good placement test has high reliability and is able to show a TTI. Cronbach and Gleser noted a difficulty with trying to apply their methods in typical educational settings. This problem was that "[t]he evaluation of outcomes, however, seems often to be arbitrary and subjective, leading one to question whether any of the conclusions from decision theory can be trustworthy if the starting point itself is open to dispute" (Cronbach and Gleser, 1957, p. 109). So the TTI technique may not provide accurate information about the placement process because grades, the payoff, are not always assigned on an objective basis.

Cronbach and Gleser's (1957, 1965) work on placement was not referenced as widely as it might have been in empirical studies of mathematics placement systems. Their theoretical position, however, was made more popular in education by Hills (1971).

Willingham (1973) characterized Hills' writing as "the most comprehensive and generally useful summary of the use of tests in selection and placement" (p. 100). Hills broadened the placement definition to be the "assignment of personnel to different treatments along a single dimension, though this may be a composite derived from a procedure such as multiple regression" (Hills, 1971, p. 701). Hills stated a rather common sense purpose of placement:

> . . . to situate the student in the course or treatment that will challenge him but will not overwhelm him - to prevent his wasting time or being bored on the one hand and to prevent his failure due to lack of preparation or lack of sufficient repetition or explication on the other (Hills, 1971, p.702).

Hills identified three criteria to evaluate the effectiveness of placement: grades, persistence, and satisfaction. He noted that there was a major problem with using each of these criteria since cutoff scores, established with variables that predict the criteria, tend to arbitrarily place students into courses. He cited the study of Dunn (1966) as a prime example. Hills discussed several methods for determining cutoff scores: quotas, probability of attainment of selected criterion level, highest-level course with selected level of predicted achievement, Cronbach and Gleser's decision theory methods, and task analysis methods of Gagné (1962). Hills did not extensively discuss this last method; however, Willingham (1974) used Gagné's theories to establish a theoretical basis for the rationale for placement systems.

Hills preferred the decision theoretic approach as a method of validating placement tests in various academic settings. He provided a more understandable discussion of each of the placement situations that Cronbach and Gleser described. Hills maintained that the fixed treatments with either fixed or adjusted quotas are the most useful for describing the placement situation in mathematics. He reiterated that in this setting a placement test is effective if there is a substantial gain in the expected payoff, where the gain is a function of the correlation between placement test scores and the underlying trait. Thus, Hills preferred to use r rather than r^2 as the appropriate effectiveness index to judge the importance of a validity coefficient.

Both the works of Cronbach and Gleser (1957, 1965) and Hills (1971) had a major impact on Willingham (1974), who was the first writer to establish placement on a firm learning theoretic basis. Willingham comprehensively discussed student placement that colleges and universities make and categorized the decisions based on the nature of a common criterion and the nature of the trait assessment (Table 1).

Table 1

Major Classes of Alternate Treatments

	Nature of the Common Criterion	
Trait Assessment	End-of-course Achievement	Other Educational Outcomes (persistence, satisfaction, etc.)
Aptitude or personal characteristic	ASSIGNMENT	SELECTION
Knowledge of subject matter	PLACEMENT	EXEMPTION

Note. Reprinted with permission from College Placement and Exemption, (p. 18), by W. W. Willingham, copyright ©1974 by College Entrance Examination Board, New York.

Willingham defined placement to be the optimal positioning of students within a sequence based on how much the student knows about the subject. He stated that the placement decision should be based on a single, perhaps composite score that predicts a common criterion measure for students who received alternate treatments. Willingham observed that the common criterion measure is frequently the final grade in the course common to both treatments. The present study is, for the most part, consistent with Willingham.

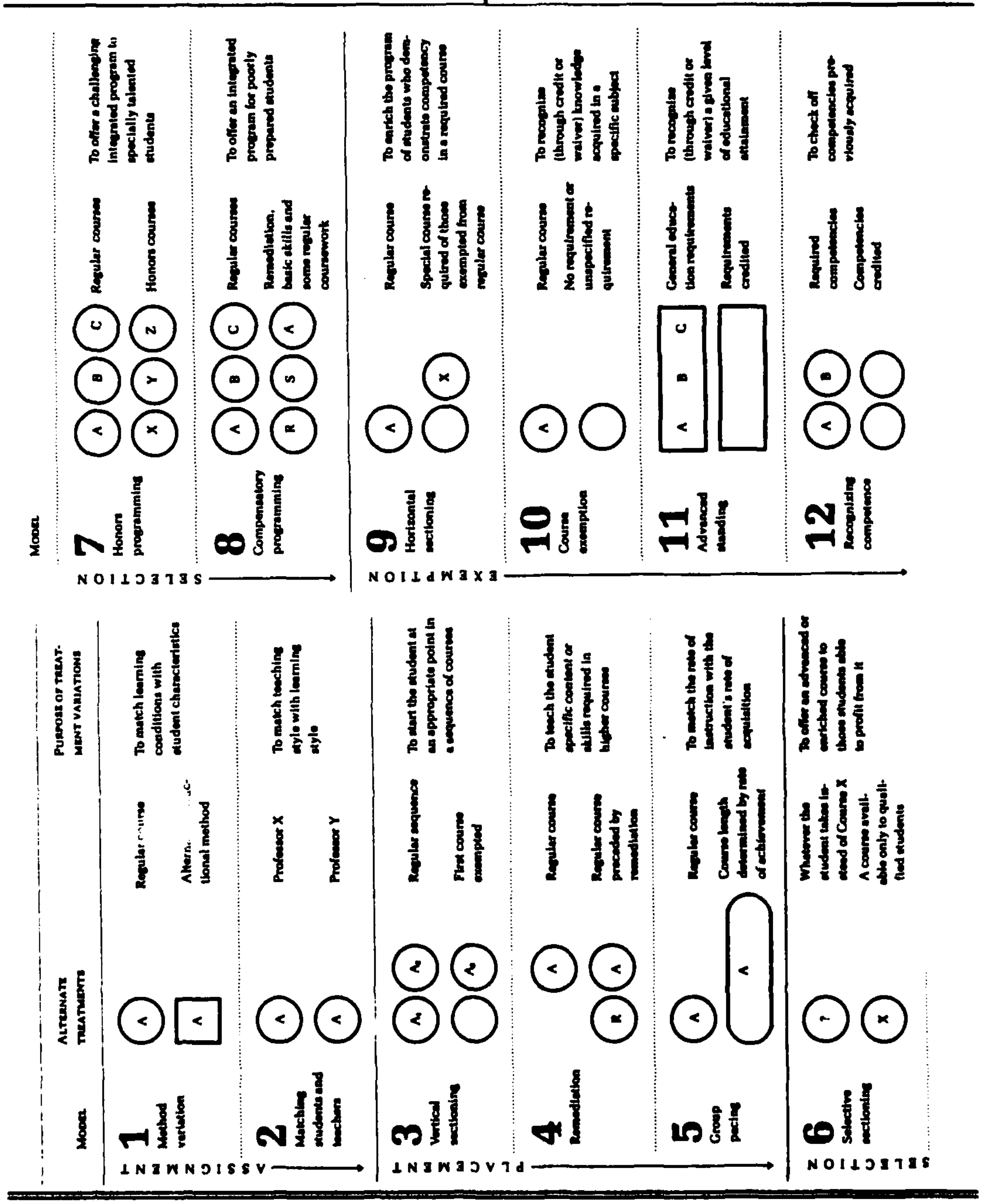
Within the framework in Table 1, Willingham developed a series of 12 alternate models (Figure 1) including the vertical placement model, model 3. Willingham used a precalculus-calculus sequence to illustrate his vertical placement model. This model called for a placement test which was closely connected to the course objectives of the precalculus course. The test functions in two ways: (a) waives the prerequisite course and (b) establishes the entry point within the mathematics course sequence.

The vertical placement model is related to the exemption models in that the exemption models are specifically used to waive course requirements. The distinction between the placement and exemption models revolves around the issue of the self-correcting and reversible nature of the decision. In vertical placement, the decision is self-correcting (i.e., the student is able to disagree with the placement) and reversible (i.e., the student may be placed into a different course). These features of the vertical placement model allow the placement test to focus on the placement issue rather than the exemption issue. The exemption test uses the comparable performance of students who satisfactorily completed the course as the basis of exemption.

The essential goal of placement is to match the capabilities of the students with the mental demands made by the courses within the course sequence. Willingham determined that the most appropriate way to accomplish this goal is to analyze the course sequence with respect to content sequence and structure. He found that Gagné's (1970) theory of learning and instruction provided a good framework to describe these processes.

Figure 1

12 Models for Exemption and Placement



Note. Reprinted with permission from College Placement and Exemption, (p. 18), by W. W. Willingham, copyright © 1974 by College Entrance Examination Board, New York.

Gagné maintained that course content should be analyzed in terms of a hierarchy of prerequisite knowledge. The knowledge may be classified into eight types of learning that range from simple signal learning to problem solving. The learning types are ordered with the simple types of learning transferring to the higher order learning. The learning hierarchy, described behaviorally, describes what needs to be measured to determine if students have acquired the capability. Willingham understood this to mean, for example, that students who are ready for calculus should be able to demonstrate the prerequisite capabilities that should have been learned in precalculus. This assessment instrument would be constructed from the behavioral objectives of the precalculus course. The students would be placed into calculus only if their scores on the placement test demonstrate sufficient mastery of the precalculus objectives.

Determining the level of sufficient mastery is an important issue for Willingham. He, like Cronbach and Gleser (1965), maintained that sufficiency of mastery, say in a precalculus-calculus sequence, is related to the alternatives. In this example, sufficiency of mastery can be translated into the likelihood of success in a short or long precalculus-calculus sequence. Thus, Willingham adopted Cronbach and Gleser's TTI techniques, with a slight adaptation, for establishing the placement test score that would maximize the success of the individual student in the alternate treatments. The adaptation was explained in the context of a placement decision involving precalculus and calculus. Willingham accepted that ". . . the best possible placement measure [of the trait] might result from weighting a, b, c [part scores on the placement test] for maximum multiple correlation with final calculus achievement for those taking the short sequence [students placed into calculus first]" (Willingham, 1974, p. 84). The criterion for an effective placement test was the presence of a TTI.

Willingham further noted, however, that the TTI method requires random placement of students of all levels of abilities into the course sequence. This is generally not feasible in the college setting since it is often ethically inappropriate to place someone with obviously deficient college algebra skills into calculus. Thus, Willingham suggested alternative ways to validate a placement examination: content validity, concurrent validity in conjunction with content validity, predictive validity, and instructional gains in a pretest-posttest setting. Willingham emphatically stated, however, that "predictive correlation between precalculus achievement and a test administered before the precalculus course does not, in and of itself, establish content validity or any sort of placement validity" (Willingham, 1974, p. 85).

Willingham further attacked the common predictive placement model and listed five shortcomings of this model:

1. Does not establish validity of the placement test.
2. Attempts to use the same placement variable at various decision levels.
3. Uses multiple correlation and stresses obtaining high $\underline{r}^2$ values rather than high correlation between the placement test and the criterion.
4. The tendency to use general ability measures which reduces the likelihood of obtaining differential placement.
5. The test is less defensible for exempting students from course requirements if it is not a fair measure of the course.

Willingham summarized his position by saying, "placement procedures must be primarily concerned with what students know, what instructional alternatives they should follow, and what outcomes of alternate treatments there are" (Willingham, 1974, p. 86). Ku and Frisbie (1978) elaborated Willingham's models of placement by describing various methods for evaluating components of a placement system.

Ku and Frisbie noted that placement systems were being implemented and operated without periodic inspection to check if the systems continued to operate effectively. They suggested that an evaluation plan be devised to monitor placement systems on a regular basis. Specifically, to validate cutoff scores they suggested evaluating the gain scores, performing follow-up studies to analyze student performance after placement, monitoring the placement and enrollment comparisons of the number of students who enroll against the suggested enrollment by the placement system, asking students about the effectiveness of their placement through student surveys, and asking for instructor feedback concerning the placement. In an appendix, Ku and Frisbie provided a table listing potential evaluation questions and associated data sources.

While Willingham developed a workable framework for placement, described its educational rationale, and provided a fairly thorough review of placement research up to that point, he did not provide enough information about how to develop and implement a good placement system. Aleamoni (1979) provided this practical informati in a systematic, eight step procedure designed to develop and implement placement systems:

1. Define the purposes of placement and exemption.
2. Determine the major instructional objectives of the course or course sequence.
3. Secure or develop an adequate test.
4. Determine the reliability and validity of the instruments through experimental administration.
5. Determine decision scores.
6. Arrange for routine administration.
7. Develop an evaluation plan.

8. Develop a procedure for periodic review and modification.

Aleamoni also highlighted the importance of determining the cutoff score and provided a 12-point procedural checklist for determining decision scores. The checklist concentrated on how to gather data and to present the suggested decision scores to decision makers. The actual method of determining the cutoff score was left to the researcher.

There is a fairly substantial body of research concerning how to determine cutoff scores. Probably the most comprehensive discussion was written by Livingston and Zieky (1982). Their manual is a thorough, yet concise description of the major procedures used to determine cutoff scores. Livingston and Zieky discussed methods based on judgments about test questions, methods based on individual test-takers, and methods based on a group of test-takers. The methods based on test questions require judges to decide if a C- student would be able to arrive at the correct solution of each test item. The methods based on individual test-takers require information about each person's test score and a judgement of the sufficiency of the person's knowledge and skills. The judges identify those persons who are minimally qualified or who were not qualified and then characterize these groups' test scores.

The last category of methods discussed were the methods based on judgments about a group of test takers. This method is consistent with Kelley's (1979) method. Kelley's method requires a reliable placement examination to be developed or selected, students to be placed into courses using some other tool, scores obtained on the placement examination items that were incorporated into the final examination, and preliminary course grades obtained before the examination scores are calculated into the official course grades. Thus, the judgments of the faculty who assign the preliminary course grades are uncontaminated by the scores on the placement examination items. Various possible cutoff scores are then determined based on:

> 1. Expected score for students whose performance in course was just minimally satisfactory -- i.e., students with preliminary grades of C- . . . .
> 2. Score for which expected grade was just minimally satisfactory -- i.e., C- . . . .
> 3. Score for which percents of errors of students in each academic performance category (Unsatisfactory, Satisfactory) were most nearly equal.
> 4. Score for which overall percents of errors were most nearly equal. . . .
> 5. Score that would have cut off . . . approximately the same number of students as were in the Unsatisfactory performance group. . . .
> 6. Score that would have maximized overall accuracy of placement. . . . (Appenzellar & Kelley, 1983, p. 15).

Recommendations for selecting the cutoff score were based on the reasonability of the recommended cutoff score, distribution of the placement examination scores, course sequence, and departmental and institutional goals.

One problem with the methods discussed by Livingston and Zieky is that the different methods have produced widely different cutoff scores over various school subjects. Crocker and Algina (1986) provided an extensive review of these standard setting comparison studies. It was clear from their review that no single method can be consistently preferred in all situations. In response to this finding, Crocker and Algina offered advice for standard setting:

1. Question whether there is a legitimate need for establishing a performance standard for interpretation of the test scores in question.
2. Identify the likely threats to validity of the inferences that are to be made from the test scores.
3. Use two or more different approaches to standard setting and multiple samples of judges.
4. Examine empirical evidence of how a typical sample of examinees perform on the test and use this information in evaluating the consequences of setting a particular standard (Crocker and Algina, 1986, pp. 419-420).

Placement Literature Summary

The literature for general placement supports the theoretical rationale of the placement systems recommended by Cronbach and Gleser (1957), Hills (1971), Willingham (1974), Aleamoni (1979), and Kelley (1979). The rationale used Gagné's (1970) theory of learning and decision theory requiring that placement goals and objectives be developed, a placement test be constructed, that a cutoff score be selected, and that techniques for assessing the placement system's effectiveness and for maintaining the system be developed. Empirical evidence was cited that suggested that several methods be used to gain a consensus to determine cutoff scores because no single method appears to give the best score in all situations.

Whereas Willingham's model focused exclusively on prerequisite knowledge, based on Gagné's theories, other theories describe learning from different points of view. Recent work in the theory of school learning has developed alternatives to describe the learning process in terms of the cognitions students should acquire to improve transfer to unfamiliar situations. The next section provides an overview of this literature used to form an alternative theoretical basis of the placement problem.

Cognitive Learning Theories

Cognitive learning theories may provide the rationale for an alternative college mathematics placement model. A general overview of several modern cognitive theories is provided, followed by specific aspects of Skemp's theory of intelligent learning and Wilson's Model of Mathematical Achievement as they apply to college mathematics placement.

Overview of Cognitive Theories

Silver (1987) reviewed much of the work in cognitive theory as it applies to mathematics. He characterized modern cognitive theories as those based on understanding of memory and information processing. In terms of memory content, most theories differentiate between "knowing that" and "knowing how". Certain cognitive science theories speak of a semantic memory that is thought to consist of concepts and relations among concepts; propositions are considered linked in memory by the associations of concepts. Other cognitive theories hold that propositions are connected in memory with whole structures representing concept relations.

Some theories involve the architecture of memory. Most theorists hypothesize that there are three kinds of memory registers: a sensory buffer, a short-term (or working) memory, and a long-term (or permanent) memory. The sensory buffer essentially receives and encodes input from the environment. Short-term memory temporarily holds and actively manipulates information. Silver (1987) reports that the short-term memory can hold only six or seven items, while long-term memory does not seem to have any storage capacity limit.

The mental activity that manipulates and controls information from the three registers is called "information processing". The advantage of efficient information processing is to reduce the cognitive strain in terms of the limits of the snort-term memory capacity. In terms of novel problem-solving, Silver noted that,

> Students' problem-solving abilities might improve greatly if they could use working memory more efficiently, that is, use automatic processing for the more routine elements of an activity, and thus make resources available for the controlled processing of the novel aspects of solving the assigned problems (Silver, 1987, p. 40).

Effective information processing in mathematical problem solving involves the ability to retrieve information from long-term memory through the efficient structure of relevant knowledge. The structure of knowledge has been pursued using the schema model of knowledge. Rumelhart and Ortony (1977) defined schema to be a cognitive data structure which represents concepts and their interrelationships with other schema. Schemas are thought to consist of variables that are embedded in other schemas with varying levels of abstraction. To Rumelhart and Ortony, schema are more than simple definitions of concepts; they represent knowledge, "the basic building blocks of the human information processing system" (Rumelhart & Ortony,1977, p. 111). Memory is composed of a set of schemas "interconnected and cross-referenced" with schemas performing at least four functions: the comprehension of sensory information, the storage of the information into memory, the retrieval of the information from long-term memory, and the making of inferences (Schallert,1982).

Important to the present study, Silver stated that schema theory provides a good explanation of why students may have difficulty solving problems which are not precisely those given by the textbook. He claimed that, "Locked into a particular approach, the student lacks the flexibility to adapt to new circumstances" (Silver, 1987, p. 48). Silver further noted that an "extensive" knowledge of mathematics was not a sufficient condition for successful problem solving; problem solving also requires a variety of "meta-cognitive" skills. Meta-cognition refers to the process whereby a student monitors, assesses, and adapts his or her current information processing activity.

In terms of learning, both Silver and Schallert identified two common principles: (a) that expertise develops with repeated experience with a variety of examples and (b) that learning is essentially a constructive activity performed by the student. Thus, schemas grow, new interrelationships develop, and new schemas come into existence only when students construct relationships between new information and their existing schemas. An instructor cannot provide the schemas ready made; students must use their own existing elements of knowledge to construct new concepts, structures, and skills.

This section presented a separate view of the concepts of information processing, constructivism, and schema in terms of learning. Skemp (1979) has developed a theory of learning which blends these individual theories into an integrated theory.

Skemp's Theory of Intelligent Learning

Skemp (1979) developed a comprehensive cognitive theory of learning and recently adapted this theory to explain learning in mathematics (Skemp, 1987). Skemp's theory integrates schema theory, information processing theory, and constructivism. This review, however, highlights the various aspects of his theory that might form the foundation for a cognitive model for college mathematics placement. Duran (1985) provided a comprehensive summary of Skemp's theory.

Skemp began by noting that much of human behavior is goal directed and that there seem to exist two cognitive systems (director systems) that systematically direct actions to achieve goal states. A director system requires (a) a sensor to take in information and represent it internally, (b) an internal representation of the goal state, (c) a comparator to compare the present state to the goal state, and (d) a plan of action of moving the operand from the present state to the goal state (Skemp, 1987).

Learning is considered a goal directed activity where the director systems change and allow for better functioning. Skemp calls intelligence "a kind of learning that results in the ability to achieve goal states in a wide variety of conditions, and by a wide variety of plans" (Skemp, 1987, p. 107). Better functioning implies increasing the domain of the director system, improving the accuracy of conceptual structures (called schemas), increasing the completeness of the schemas, and improving the skill of actualizing the plans (Skemp, 1979).

Skemp views learning as the building, testing, and maintaining of conceptual structures. Skemp described two director systems that perform separate parts of these activities. The delta-one director system operates on environmental input, while the delta-two director system operates on the delta-one director system. The goal of delta-two is to improve the functioning of delta-one. An important goal of delta-two is to construct schema. Delta-two constructs schema in order for delta-one to function better within the environment and derive plans from the schema so that delta-one will be able to operate effectively in more varied circumstances. This is what Skemp terms "intelligent learning" (Skemp, 1979, p. 85).

According to Skemp, schemas have several characteristics that are useful in describing learning in mathematics:

(i) A schema is a structure of connected concepts. . . . A schema in its general form contains many levels of abstraction, concepts with interiority, and represents possible states (conceivable states) as well as actual states.

Factors important for effectiveness of schema include:

(ii) Relevance of content to the task in hand.
(iii) The extent of its domain.
(iv) The accuracy with which it represents actuality.
(v) The completeness with which it represents actuality within this domain.
(vi) The quality of organization which makes it possible to use concepts of lower or higher order as required, and to interchange concepts and schemas.
(vii) By a high-order schema, we mean one containing high-order concepts.
(viii) The strength of the connections.
(ix) The quality of the connections, whether associative [concepts memorized without connections] or conceptual [connected to appropriate schema].
(x) The content of ready-to-hand plans which remain integrated with the parent schema.
(xi) Penetration - the degree to which it can function in high-noise conditions.
(xii) Assimilatory power - the degree to which it can assimilate new experience.
(xiii) Assimilatory power relative to other schemas . . . (Skemp, 1979, pp. 190-191).

Skemp noted two types of understanding in school mathematics: instrumental and relational. Instrumental understanding is defined to be memorized rules without meaning -- for example, in differential calculus, memorized rules of differentiation without understanding their relationship to the definition of the derivative. Relational understanding means knowing both the how and why in mathematics; this could mean knowing the relationship between the rules for anti-differentiation and limit theory. In the first case, there are few conceptual links; most of the links are associative, which virtually isolates the concept. Skemp (1987) reported that this kind of understanding has been shown to be effective in the short term but is disadvantageous in the long term: (a) instrumentally learned mathematics is usually easier to understand; (b) the rewards are more immediate, and more apparent; and (c) one can often get the right answer more quickly and reliably when problems are proposed exactly as learned. Skemp (1979) further suggested that the advantages of relational understanding far outweigh those for instrumental understanding. The advantages of relational understanding are: (a) it is more adaptable to new tasks; (b) it is easier to remember; (c) it may be a goal in itself; and (d) relational schemas are organic, in that they grow from within, without direct instruction.

So, relational mathematics knowledge is more adaptable to growth and can be used in novel settings more effectively and appropriately than can instrumental understanding. Thus, Skemp's theory predicts that students with relational schemas will be more likely to succeed in learning subsequent mathematics than students with instrumental schemas.

Skemp delineated the problems that the instrumental learner will eventually face:

> The problem here is that a bright and willing child can memorize so many of the processes of elementary mathematics so well that it is difficult to distinguish it from learning based on comprehension. Sooner or later, however, this must come to grief, for two reasons. The first is that as mathematics becomes more advanced and more complex, the number of different routines to be memorized imposes an impossible burden on the memory. Second, a routine only works for a limited range of problems and cannot be adapted by the learner to other problems, apparently different but based on the same mathematical ideas (Skemp, 1987, p.94).

Hence, students with relational prerequisite mathematical schemas are more likely to succeed than those with instrumental schemas. One significant problem arises in trying to apply this to the placement problem, especially if placement tests are constructed according to Braswell's criteria (cited in MAA, 1983), "In general, questions that are nonroutine or insightful in nature are not appropriate for use on placement tests" (p. 6). Thus, according to Braswell's approach the entire test should be composed of the type of items which would favor students with instrumental mathematical schemas! Clearly, a methodology must be developed to assess the degree of students' relational schemas to be consistent with Skemp's theories.

Greeno (1978) suggested three criteria for evaluating the degree of understanding within a semantic network: (1) internal coherence, or completeness, of the representation; (2) connectedness of the information to other schemas; and (3) correspondence, or accuracy, of the representation.

Resnick and Ford claimed that the best way to assess internal integration is by measuring the access time and noting the pattern in which a number of items are related when answering a question, e.g., "Tell me everything you know about Newton's law of moments" (Resnick & Ford, 1981, p.207). Resnick and Ford further suggested that diagnostic interviews should be used to assess the degree of connectedness. They also recommended assessing the correspondence of representation by the association method, where "a person is given a word drawn from the subject matter domain and is asked to state as many other words or concepts associated with the target word as possible. Once a list of associated words is collected for each target word, it is possible to compare the lists [the student's and an expert's] for degree of overlap" (Resnick and Ford, p. 208). Each of these situations seem impractical in the typical college placement setting. While Skemp (1987) promoted using diagnostic interviews to assess the quality of schemas, he also suggested an alternative assessment method: Measuring the ability to solve novel problems. This method is based on the assumption that relationally understood schemas are more adaptable to unfamiliar situations.

He agreed with Backhouse who stated that,

> We are unable to observe our pupils' feelings and schemas directly, and look for confirmatory behaviour. As evidence that A understands X [relationally], we accept the fact that A applies X in situations different (in greater or less degree) from that in which it was learned (cited in Skemp, 1987, pp. 166).

Skemp's theory, as it applies to mathematics placement, may be summarized with the following principles: (a) students must possess the necessary prerequisite relational knowledge to enjoy long-term success in learning mathematics, (b) students with relational schemas are more likely to succeed in learning subsequent mathematics than students with instrumental schemas, and (c) the ability to apply prerequisite knowledge in novel settings distinguishes students with relational or instrumental schemas.

Thus, Skemp's theory supplies a basic rationale for understanding the placement problem by defining a set of behaviors that may be used to infer the cognitive conditions in students and linking them to the successfulness in learning subsequent mathematics. This rationale can be used to develop a placement test prepared from a conceptual analysis of the prerequisite course. Placement test items should require students to apply prerequisite knowledge in novel settings in order to measure the degree of connectedness and internal integration of the prerequisite schemas. Other items also should be included to assess the degree to which the prerequisite knowledge is possessed by the learner.

This investigator argues that two subscales, one composed of novel items and the other composed of items to assess completeness of the prerequisite schemas, can be used to predict the success of students in learning subsequent mathematics. Scores on this type of examination should predict success. In other words, the examination should be useful for placing students into mathematics sequences.

Skemp's theory of learning suggests the characteristics of a valid placement test, but his theory does not supply the specific technology that can be used to write the necessary test items. Skemp's theory does not give operational definitions for a novel problem or a method for distinguishing between novel and routine problems. Wilson's (1971) Model of Mathematics Achievement provides a framework for making these decisions.

Wilson's Model of Mathematics Achievement

Wilson (1971) supplied a detailed analysis of evaluation of mathematics learning. His model is a two-way content-by-behavior (cognitive and affective) taxonomy and represents an expanded version of the model of mathematical achievement used in the

National Longitudinal Study of Mathematics Achievement (NLSMA) (Romberg & Wilson, 1969). The Wilson model is consistent with the model established by Avital and Shettleworth (1968) and represents a synthesis from Bloom's (1956) taxonomy of educational objectives.

The four basic levels of cognitive mathematics behavior in the model reflect the cognitive complexity generally required in secondary school mathematics. These levels are computation, comprehension, application, and analysis. Wilson supplied an example of his model for secondary school mathematics, grades 7 through 12. Within that context, he provided finer distinctions for each basic level of cognitive behavior, as well as affective behavior. He maintained that the basic levels of cognitive behavior are both hierarchical and ordered. The content dimensions Wilson used represented the major content divisions, with additional subdivisions within each major area.

The primary importance of Wilson's model to the present study is his fourth level of cognitive behavior. Skemp and Backhouse advised that it is possible to assess relational understanding with items that require students to apply schema in novel settings. In terms of the Wilson model this means that placement test items should be written for specific college mathematics content areas at the analysis level of cognitive behavior, as that level primarily represents non-routine problem-solving. Wilson provided the necessary operational contrast between analysis and non-analysis level items. He observed that in all cases analysis level items differ from others in that they involve ". . . a degree of transfer to a context in which there has been no practice. . . . the student is given a problem situation for which an algorithmic solution is not available to him" (Wilson, 1971, p. 662).

In Skemp's terms the possession of a relational schema for a particular mathematical concept implies that the learner possesses the capability to generate not only the definition of the concept but also various representations of the concept and its connections to other concepts with a high degree of fluency. On the other hand, Wilson's classification of a test item at the analysis level of cognitive behavior implies that the correct response to the item is not available by recall from memory but instead requires a new organization of related concepts. One infers that such a reorganization is possible only if the essential concepts and connections are either possessed by or can be generated by the student.

Wilson's model augments the theoretical position established by Skemp by providing a content-by-cognitive behavior framework to classify various levels of mathematical behaviors which result from a conceptual analysis of the prerequisite course, and by providing the operational definitions for constructing or identifying novel problems that may be used to assess these mathematical behaviors. Together Skemp's theory and Wilson's model provide the major elements of a cognitive model for placement within college mathematics.

The Cognitive Model of College Mathematics Placement

Question I, presented in the Introduction, asked if a cognitive model for college mathematics placement could be developed from Skemp's learning theories and Wilson's model of mathematics achievement. This section is devoted to answering this question by describing such a model and discussing ideal and alternative methods for validating placement examinations based on the model.

The model assumes that the placement problem is to maximize the likelihood of success of students within a college mathematics sequence involving a prerequisite and a criterion course, i.e., the vertical placement problem. Students are most likely to succeed when their prerequisite mathematical schemas (a) are well connected with a high degree of internal integration, and (b) have a high degree of correspondence and completeness with the selected prerequisite schemas necessary for success in the follow-on course. The degree of the connectedness and internal integration of the students' prerequisite mathematical schemas may be measured with placement test items written at the analysis level of cognitive behavior. The degree of correspondence with selected prerequisite schemas may be assessed with other items written at the computation, comprehension, and application levels of cognitive behavior.

A placement test developed using the model should contain two subscales. One subscale should be composed of analysis level items relating the major content divisions represented in the prerequisite course. These items should not be overly difficult or tricky; rather they should require flexible and fluid mathematical thought to identify relational schemas. The other subscale should be composed of non-analysis level items thoroughly covering the variety of the content divisions of the prerequisite course identified as necessary for success in the criterion course; the focus of this subscale is on the completeness of the students' schemas. A weighted composite of the students' scores on these two subscales may be used to identify the students most likely to succeed in the criterion course; these students should be placed directly into the criterion course. Those students who are identified as unlikely to succeed in the criterion course should be placed into the prerequisite course.

Ideally, the composite score which differentiates the potentially successful and unsuccessful students, should be determined through the use of TTI techniques. Initial estimation of the appropriate cutoff score requires that placement subscale scores be obtained from the entering freshman students before they are placed into any mathematics courses. The students should then be randomly placed into the prerequisite and criterion courses and allowed to complete the sequence. TTI techniques should be applied to the students' final grades in the criterion course, placement sequence, and placement subscales scores to (a) validate that differential prediction exists, and (b) to estimate the weighted composite score that will indicate most accurately the students' likelihood of success in the criterion mathematics course. The results should be cross-validated, preferably with a different group of individuals.

Practically, the TTI technique described above can rarely be used since the collegiate placement setting usually does not allow for randomly placing students into courses. In this case course placement should be performed as accurately as possible and alternative methods for validating the prospective placement test should be used. The content and construct validity of the placement test become the major focus of concern. A high degree of agreement among judges that the items are representative of content problems in the prerequisite course should indicate that the subscales adequately cover the content of the prerequisite course. The construct validity should substantiate the composition of the subscales in terms of analysis and non-analysis level items. In addition, the predictive and concurrent validity of the placement test may also be investigated. Predictive and concurrent validity coefficients in the range of 0.40 -- 0.60 should be expected. Most existing placement test procedures could be converted to the Cognitive model placement procedure through the division of the placement examination total score into its analysis and non-analysis subscale components.

When a college is unable to develop non-analysis subscales which thoroughly cover the content of the prerequisite course; other assessments of the completeness of the prerequisite schema may be needed. In this situation a supplemental estimate of the two measures may enhance the accuracy of the placement. The next section reviews such measures.

Placement Variables

The Cognitive Model for College Mathematics Placement provides criteria to evaluate variables to be used to predict successful placements within a college mathematics sequence. Specifically required are prediction variables which (a) assess the quality of students' prerequisite relational schemas and (b) assess the completeness of the students' prerequisite schemas. This section evaluates the variables commonly used in empirical placement studies against these criteria.

Final Grades

Final grades at the end of an instructional sequence are by far the most common predicted or criterion variable used in placement studies. Final grades are generally thought to reflect student achievement in the course content; however, this may not always be true, because final grades in college courses may be contaminated by numerous other influences (Cronbach & Gleser, 1957; Hills, 1971; McComb, 1987). McComb observed two major contaminants: (a) grades may not represent achievement in the subject matter (e.g., grades also may reflect for attendance, extra credit work, class participation, etc) and (b) grades may not be assigned consistently across instructors or offerings (e.g., class participation may not be assessed consistently, grading "on a curve", etc.).

Given these weaknesses it is apparent that alternatives should be sought. One alternative that Willingham (1974) and Appenzellar and Kelley (1983) suggested is an objectively scored, comprehensive common final examination. Scores from this type of examination may resist the inherent problems with final grades, although this measure may have different threats to reliability and validity. These threats may come from teaching effects (e.g., not all the syllabus was covered) or personnel effects (e.g., was feeling ill the day of the examination).

Placement Examinations

A locally or commercially developed placement test is, perhaps, the most common predictor variable in previous placement studies. Such predictors have been incorporated into decision theoretic models (Cronbach & Gleser, 1957; Hills, 1971; Willingham, 1974; McComb, 1987) as well as empirical prediction models (Bingham, 1972; Bridgeman, 1980; Appenzellar & Kelley, 1983; Eshenroder, 1987).

When TTI-based models are used to investigate placement procedures, it is expected that the correlation between placement test scores and final grades in the criterion course should be lower for students in the long sequence than for students in the short sequence when the prerequisite course is operating effectively. This means, presumably, that the students who did not initially possess the prerequisite capabilities, as measured by the placement tests, acquired them in the prerequisite course. A low correlation here indicates that, in the long sequence, the students' criterion performances are less dependent on the level of achievement represented by the placement examination scores. Medium to high correlations between placement scores and final grades in the criterion course for students in the long sequence are presumed to indicate that the prerequisite course was not very effective as a preparatory course.

None of the placement tests used in the above mentioned empirical studies were constructed according to the Cognitive model. None of the studies described the cognitive structure of the test they used. In fact, very little was reported concerning the development of the tests besides stating the mathematics content they covered together with estimates of reliability. The studies based on decision theory or on Willingham's model were more likely to report other information related to the validity of the tests, while virtually all of the studies reported the predictive validity of the tests.

A properly constructed placement examination is essential for establishing a TTI between alternative sequences and for implementing effective placement. Perhaps existing examinations could be analyzed to form the necessary subscores consistent with the Cognitive model. Locally developed examinations would be more accessible to this analysis than commercially developed placement tests, since commercial test developers tend to restrict the necessary test item data for test security reasons.

Other Predictor Variables

Several other predictors are commonly used in empirical placement studies. Results from a Mathematics Association of America (MAA) Placement Test Program (PTP) questionnaire (Cederberg & Harvey, 1987) showed that 85% of a sample of 84 PTP subscribers reported using some measure other than a placement test to perform their college mathematics placement. Measures of general mathematics ability or mathematics achievement, nationally normed, are frequently used. Two of these predictors are scores from The College Board Scholastic Aptitude Test, Mathematical Score (SAT-M) and the American College Testing Program Mathematics Score (ACT-M) subtest. The SAT-M may function differently from the ACT-M within the Cognitive model context because the SAT-M is not closely tied to any specific college preparatory curriculum (Angoff, 1971). The ACT-M has traditionally been based on typical mid-western high school curricula. Since the context of the test item affects its cognitive classification, items written without a particular curriculum in mind logically seem to enhance the degree of novelty to the test takers.

While it is senseless to discuss measuring interrelationships between concepts if the concepts have not been learned, it is assumed that most college preparatory curricula include the basic concepts measured by the SAT-M. Even though the students may not be "schooled" on the solution of specific item types, the students should be able to solve the problems if their schemas are flexible and adaptable. A serious difficulty with this line of thought is that without knowing the items and their cognitive classifications, it is not possible to determine the extent to which the scores contribute to the assessment of the students' relational schemas. Indeed, the SAT-M or ACT-M tests scores should be judged as inappropriate to be used as measures of the extent of students' relational schemas for precalculus without their associated cognitive subscale scores.

SAT-M and ACT-M examination scores may also be evaluated in terms of their contributions to the assessment of the completeness of the prerequisite schemas. Both examinations seem to provide reliable, although general, measures of certain prerequisite schemas required for freshmen entering college mathematics sequence (Angoff & Dyer, 1971). The validity of these assumptions may be reflected in the correlations between the SAT-M scores and the final end-of-first semester mathematics grades in the criterion course. Historically, these correlations range between .40 and .70 for large samples .

Other prerequisite mathematical schemas may be necessary for success in the calculus that are not assessed by the SAT-M or the ACT-M. For example, Skemp (1987) maintained that symbolic and logical schemas are also often needed. These schemas relate to the ability to connect mathematical symbolism and notation with relevant mathematical ideas, and to cope with mathematical rigor. It is conceivable that the possession of these and other schemas may be required for success in the criterion course and yet are not addressed by the objectives of the prerequisite course. If these schemas are not present and not taught in the prerequisite course, then neither sequence will be effective. Therefore, it is possible for a student to possess some of the prerequisite schemas, supported by the prerequisite course, and yet still not be ensured of success in the criterion course. This situation may be alleviated in at least three ways by providing instruction of these schemas in: (a) the prerequisite course; (b) the criterion course; or (c) in another concurrent course.

Other writers have suggested that many noncognitive schemas (e.g., persistence, study habits, favorite subject, motivation) and individual characteristics (e.g., age, sex, date of admission) relate to the success in many college mathematics courses (Tinto, 1982; Owens, 1987; Eshenroder, 1987). While the proposed model of placement has focused on cognitive factors, theories upon which it is based are not so restricted; the model could be expanded to include "affective" factors as well. However, it seems that while affective factors previously have been shown to be statistically related to success, it has been difficult to explain or attribute causality to the relationship. Not enough is known about these variables and criterion relationships for them to be included in the Cognitive model at this time.

Another frequently used variable in placement studies is the students' prior academic performance (Hunt, 1987; McKillip, 1966; Morgan, 1970; Owens, 1986; Wick, 1965). Ahrens (1980) used several such variables, including high school background, as measured by the number of credits taken in high school mathematics, and the course recommendation from the student's advisor. Eshenroder (1987) investigated high school grade point average (GPA), number of mathematics credit hours enrolled in the previous quarter, date of latest mathematics course, grade in the last high school mathematics course, years of high school mathematics, and the number of high school mathematics courses completed. The general rationale for including these variables is that they were used in previous studies with some degree of predictive success, that they were available, and that they logically relate to the success of students in college mathematics courses. However, according to the Cognitive model few, if any, of these variables can be tied theoretically to the prerequisite schemas of concern in the placement problem as described so far. They may relate more to general affective or social mathematics characteristics of students, so they will not be considered further as variables for inclusion in the Cognitive model.

All variables considered for inclusions in a placement model should be judged in terms of stability across time in large samples. This researcher found that studies that have included cross-validation of the predictors have used the Willingham model (Bridgeman, 1980, and Appenzellar & Kelley, 1983) with mixed results. Most other studies involved only one sample, oftentimes using small numbers of observations. A tacit assumption of the Cognitive model is that the analysis and non-analysis subscores achieved on a properly constructed placement test are stable predictors of student success.

Summary of Philosophy for Selecting Placement Variables

The basic philosophy for selecting placement variables in this research study was to judge them in terms of the Cognitive Model for College Mathematics Placement. The variables must have been able either to assess the completeness or correspondence of the prerequisite schemas for the criterion course, or to assess the quantity and quality of the connections between the prerequisite concepts necessary for the cultivation of the schemas in the criterion course. Locally developed placement tests should be the best source of these measures. These tests should provide subscores for both types of assessments when broken down into cognitive level subscales, analysis and non-analysis level items. Additionally, SAT-M and the ACT-M scores were argued to provide a reasonably thorough measure of the completeness of some of the schemas necessary for most calculus courses. Finally, placement variables should be further validated for stability through cross-validation, using large samples.

Summary of Review of Previous Research

This chapter reviewed the major theoretical bases for placement models--decision theory and Willingham's models. Willingham's Vertical Placement Model was based on Gagne's theory of instruction, which emphasized the necessity of students possessing prerequisite knowledge to promote subsequent mathematics learning. Skemp's learning theory also placed an emphasis on prerequisite schemas, but he went on to state that students who possessed many conceptual links between prerequisite schemas were more likely to be successful in learning mathematics than those who did not. Skemp further argued that novel problems could assess the degree to which students' schemas were flexible and adaptive. Wilson's Model of Mathematics Achievement developed a content-by-cognitive behavior taxonomy to classify mathematics achievement. His analysis level of cognitive behavior was shown to be particularly useful to this research as this type of items corresponds to Skemp's novel problem.

The Cognitive Model for College Mathematics Placement was proposed, based on a synthesis of Skemp's and Wilson's theories. The major theoretical assumption of the model is that measures of the degree of students' completeness of the prerequisite schemas, and the degree of the adaptability and flexibility of their schemas, may be used to predict their success in subsequent mathematics learning.

The model also contained ideal and practical suggestions for validating tests constructed or adapted using the Cognitive Model.

Also included in this section was a review of the various predictors of final grades that have been used in previous mathematics placement and prediction studies. The variables determined to be consistent with the Cognitive model were placement test scores, classified by cognitive levels, and the SAT-M and ACT-M scores. The SAT-M and ACT-M scores were assumed to estimate the completeness and correspondence of the students' schemas with some of the prerequisite schemas required by a differential calculus course. Non-analysis items on the placement tests also estimate the completeness of prerequisite schemas. Thus, scores on analysis items on the placement tests, scores on the non-analysis items on the placement tests, and SAT or ACT mathematics scores should demonstrate a TTI with long and short precalculus-calculus sequences under appropriate sampling conditions.

Other noncognitive (affective and social) variables were not considered in the present model; however, both Skemp's theory of mathematics learning and Wilson's model contain noncognitive elements. Therefore, the Cognitive Model for College Mathematics Placement may be expanded to accommodate a wider collection of variables for use in college mathematics placement at a later time.

CHAPTER III

METHODOLOGY

This chapter describes the design, procedures, subjects, placement environment, and the limitations of the study. Prior to this description, however, is a list of operational definitions which supplement those presented in Chapter I.

Definitions

The following operational definitions were used by this investigator:

Academic Composite Index (ACI): A weighted average of the prior academic record (PAR), ACT-E, ACT-M, ACT-N, ACT-S, SAT-M, SAT-V scores used by USAFA for admissions and placement. The ACI ranges from 2,000 to 4,000.

Hand-placed Cadets: Cadets who were actually placed in a course sequence different than that specified by a placement model.

Hypothetical placement: The conjectured placement of students into course sequences by a placement procedure other than what was actually used.

Generalized *F*-test: A statistical test that compares the fit of two regression equations to the same data. In this research, the generalized *F*-tests used the following formula:

$$F^* = \frac{(RSS_i - RSS_f)/(DF_i - DF_f)}{RSS_f/DF_f} \qquad (1).$$

RSS_f and DF_f are the residual sum of squares and the degrees of freedom from the Cognitive model, and RSS_i and DF_i are the residual sum of squares and the degrees of freedom from an alternative placement model. The calculated $\underline{F}^*$ value was compared to the $\underline{F}(\alpha = 0.05, DF_i - DF_f, DF_f)$ tabled value to test for significance. An observed $\underline{F}^*$ value greater than the tabled $\underline{F}$ value was interpreted as indicating that the Cognitive prediction equation explains more of the variance in the data than the alternative prediction equation.

Kuder-Richardson Formula-20 (KR-20): A measure of reliability (internal consistency). The formula is:

$$\text{KR-20} = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} p_i(1-p_i)}{\hat{\sigma}}\right)$$

where k is the number of items on the test, p_i is the proportion of students who answered the i^{th} item correct (assuming right-wrong scoring), and $\hat{\sigma}$ is the total test standard deviation.

Prior Academic Record (PAR): A score based on either the cadets' percentile rank in their high school graduation classes, adjusted for the percentage of graduates who continue on to college, or an equivalent score based on their high school grade point averages, adjusted for honors or Advanced Placement courses. The PAR ranges from 800 to 200.

Shrinkage: An index used to cross-validate prediction equations. Shrinkage values are measures of the stability of the accuracy of the prediction equations over time.

USAFA and DFMS Background Information

The U. S. Air Force Academy is a military service academy awarding the standard bachelor of science degree recognized by the North Central Association of Colleges and Schools. The mission of the Academy is to prepare the future leaders of the Air Force. The program at USAFA includes military studies, aviation experience, military training, physical fitness training, and personal standards training as well as academics. Upon graduation, the students are commissioned as junior officers in one of the uniformed services. USAFA faculty are primarily active duty military officers who serve as academic instructors and officer role models.

The Department of Mathematical Sciences (DFMS) supports the mission at the Academy through providing instruction for a common core of mathematics courses which all cadets must successfully complete. DFMS also provides other upper division mathematics service courses for specific majors as well as providing courses for its own mathematics and operations research majors. The DFMS faculty is composed of approximately 52 officers, all of whom have earned at least a Masters of Science degree in mathematics or some technically related field. Approximately 20% of the faculty have also earned PhD or EdD degrees.

Two course sequences were studied with historical data obtained in the fall semesters of 1985, 1986, and 1987 from the cadets in the classes that will graduate in the years 1989, 1990, and 1991. The long sequence was composed of two courses, precalculus and calculus I (differential calculus). These two courses were offered in each of the fall semesters of 1985 through 1987 with the precalculus course considered as a prerequisite to calculus. Precalculus was a one-half semester course which met every day and covered college algebra and trigonometry. The course was offered only in the first half of the first semester of the cadets' freshman year. The second course in the long sequence was a half-semester calculus course which met every day and covered functions, conic sections, limit theory, differentiation, and applications of differentiation. This course was only offered in the second half of the first semester of the cadets' freshman year.

The other sequence was identified as the short sequence because it was composed of a single differential calculus course. This course was a full semester course offered in the first and second semesters of the cadets' freshman year. The course used the same textbook and covered the same content sections as the calculus course in the long sequence.

All freshman core mathematics courses, including precalculus and calculus, were managed by one person, the division chief, who supervised the course directors. The course directors were directly responsible for their course materials, course interface with the other inter- and intra-departmental courses, and course evaluation. The course directors prepared the tests, syllabus, cadet and instructor notes, and the course-wide plan for grading the tests. They also supervised the team grading of the tests to ensure consistent scoring. At the end of the course, they assigned grades to the cadets on a course-wide basis. The cadets' grades in different sections were computed strictly from the composite scores on the common quizzes, common midterm examinations, and the common final examination. Cadets' grades for a given course were comparable between offerings as the division chief ensured consistency between semesters within an academic year. Additionally, final grades were not officially assigned until approved by the DFMS department chairman and the Dean of the Faculty.

Minor adjustments were made in the syllabi, instructors' and cadets' notes, examinations, and grading policies for the Classes of 1989 through 1991. For example, a new textbook was selected for the precalculus course for classes after 1989. The text was an updated edition of the text already being used, which caused slight changes in the course but did not impact the precalculus placement tests for the Classes of 1990 and 1991.

DFMS Placement Procedures

Soon after arriving at the Academy, cadets completed a battery of mathematics placement examinations administered by DFMS personnel. They were then placed into either precalculus, Calculus I, or some other core mathematics course as discussed below. Cadets could challenge their placement. If the challenges were approved, the cadets were put into courses of their choice. The DFMS placement officer evaluated the petitions and changed placements on a case-by-case basis. The number of approved challenges were 41, 11, and 5 for the Classes of 1989, 1990, and 1991 respectively.

All the placement data used in the study were obtained from DFMS sources in the fall of 1988. Thus, the cadets had already taken the placement examinations, had been placed into their initial courses, and had received their final calculus grades. These three classes of cadets were placed using one of two empirical placement systems. The placement decisions for the cadets in the Classes of 1989 and 1990 were made according to DFMS Standard Operating Procedure, SOP P-2 (see Appendix A). This operating procedure described in detail the method for placing all of the freshmen cadets into the core mathematics sequence, including precalculus and calculus. The placement decision for precalculus and calculus was based primarily on the algebra placement test total score; however, other factors were also considered. These other factors were: (a) the trigonometry and calculus placement tests total scores, (b) SAT-M or ACT-M scores, (c) previous enrollment in equivalent courses, (d) USAFA Preparatory School experiences, (e) ACI, and (f) SAT-V or ACT-E scores. For certain ranges of the algebra placement test total score, "very positive" values for the other variables, or "something negative", influenced borderline cases. This procedure will be referred to as the SOP placement model.

The Class of 1991 was placed using the Computerized Placement Model (Boudet, 1987). This model was an empirical placement model based on linear discriminant function analysis and was designed to place all freshman cadets throughout the entire core mathematics sequence, not just the precalculus and calculus courses. Different discriminant functions were used for placing USAFA Preparatory School graduates and non-Preparatory School graduates. In addition, non-Preparatory School graduate cadets who scored below 45% on the algebra placement test were automatically placed into precalculus, while no USAFA Preparatory School graduate was allowed to be placed into precalculus. The predictors for the non-Preparatory School graduates who scored above 45% on the algebra placement test, in the order of importance, were the: Calculus II placement test score (CALC2), algebra placement test score (ALG), Academic Composite Index (ACI), Calculus I placement test score (CALC1), trigonometry placement test score (TRIG), and Calculus III placement test score (CALC3). The variables used to place USAFA Preparatory School graduates were essentially the same but were found to be in a different order of importance. The trigonometry placement test score was found to not be a significant predictor of success for this last group.

Subjects

The study used admissions and mathematics placement data from the freshman cadets of the United States Air Force Academy (USAFA) of the classes that will graduate in the years 1989, 1990, and 1991. These groups were called the Classes of 1989, 1990, and 1991 respectively. The backgrounds of these classes were remarkably similar in terms of the distributions of gender, ethnicity, and many cognitive measures. The cadets' characteristics are described in Table 2 by class. In general, the classes were predominantly non-minority males with relatively high verbal and mathematical capabilities.

Since the research focused on placement within the precalculus-calculus sequences, not all freshman cadets were involved in the study. The cadets involved in this study were those who had placement examination scores, placement examination item data, and who either were placed directly into calculus and received a grade, or were placed first into precalculus and received a grade and then subsequently completed calculus and received a grade. The number of students in the study is given in Table 3.

Table 3
Number and Percentage of Cadets by Sequence and Class

	Long Sequence		Short Sequence		Total	
Year	n	%[a]	n	%	n	%
1989	121	8.8	461	33.5	582	42.3
1990	123	9.2	569	42.8	692	52.0
1991	185	13.7	557	41.3	742	55.0

Note. [a] % represents the percentage of the class in this sequence.

Variables

The variables used in this study were obtained from the DFMS cadet placement files, were constructed from DFMS mathematics placement files, or were obtained by a validity questionnaire (see Appendix B). These variables were the Academic Composite Index (ACI), the equated SAT-M -- ACT-M scores (MATH), algebra placement test score (ALG), algebra analysis subscale score (ALGA), algebra non-analysis subscale score (ALGNA), trigonometry placement test

Table 2
Entering Cadet Characteristics

CATEGORY	1989 X̄	1989 (n)	1990 X̄	1990 (n)	1991 X̄	1991 (n)
Entered						
Men		(1199)		(1178)		(1145)
Women		(176)		(152)		(203)
Total		(1375)		(1330)		(1348)
Minority						
Total		(196)		(174)		(202)
ACI						
Men	3133	(1199)	3172	(1178)	3174	(1145)
Women	3169	(176)	3212	(152)	3218	(203)
Total	3137	(1375)	3177	(1330)	3181	(1348)
Mean ACT ENGLISH						
Men	24.3	(627)	24.4	(585)	24.5	(548)
Women	25.4	(101)	25.1	(92)	25.2	(100)
Total	24.5	(728)	24.5	(677)	24.6	(648)
Mean ACT-M						
Men	29.7	(627)	29.6	(585)	29.5	(548)
Women	28.8	(101)	28.2	(92)	28.5	(100)
Total	29.6	(728)	29.4	(677)	29.3	(648)
Mean SAT-V						
Men	572	(572)	575	(593)	579	(597)
Women	597	(75)	589	(60)	583	(103)
Total	575	(647)	576	(653)	579	(700)
Mean SAT-M						
Men	658	(572)	663	(593)	668	(597)
Women	638	(75)	658	(60)	647	(103)
Total	656	(647)	663	(653)	665	(700)
PAR						
Men	629	(1199)	645	(1178)	642	(1145)
Women	661	(176)	681	(152)	681	(203)
Total	633	(1375)	649	(1330)	648	(1348)

score (TRIG), trigonometry analysis subscale score (TRIGA), trigonometry non-analysis subscale score (TRIGNA), Calculus I placement test score (CALC1), Calculus II placement test score (CALC2), Calculus III placement test score (CALC3), and the final grade in Calculus I (GRADE).

Dependent Variable

The final Calculus I grade (GRADE) was the dependent variable in the prediction equations. As previously described, this variable was relatively uncontaminated with individual instructor grading biases, subjectively scored tests, inconsistent syllabi across sections, differences in textbooks, and differences in content and structure of tests.

Independent Variables

The independent variables of prime interest in the study were the algebra and trigonometry placement examinations scores covering the content presented in the precalculus course. These tests, as well as the other placement tests, were developed by DFMS personnel at USAFA. The algebra placement test was a 40-item, five choice, multiple-choice formatted test. The trigonometry placement test was similarly formatted but contained 20 items. The same forms of these examinations were administered under similar conditions to all cadets in the three year groups. The calculus placement examinations, producing CALC1, CALC2, and CALC3, were similarly formatted; however, mostly volunteers took these tests.

The two other independent variables, ACI and MATH, were both constructed from a variety of data. The variable ACI was constructed by USAFA and was used both for admission to USAFA and for placement in mathematics. The index has been computed using the following three formulae since 1977. For cadets who took the SAT examination:

$$ACI = 2.07*PAR + 1.0*SAT\text{-}V + 1.99*SAT\text{-}M - 48.$$

For cadets who took the ACT examination:

$$ACI = 2.02*PAR + 10.4*ACT\text{-}E + 37.1*ACT\text{-}M + 7.67*ACT\text{-}S + 15.7*ACT\text{-}N - 188.$$

For USAFA Preparatory School graduates:

$$ACI = 688*(\text{Preparatory School cumulative GPA}) + 811.$$

The variable MATH was constructed for this study. Each cadet in the study had a SAT-M or ACT-M score, but no person had both scores. This investigator previously argued that these two scores measure the completeness of a part of the prerequisite precalculus knowledge for a typical graduated high school senior. The scores were equated using the equipercentile equating tables developed by Langston and Watkins (1980). The converted ACT-M scores were computed by taking the midpoint of the corresponding class of equipercentile SAT-M scores. For example, the ACT-M score of 35 corresponded to the class of equipercentile SAT-M scores of 740 - 750, with 745 as its midpoint. Thus, in this example, the converted ACT-M score was 745. The converted ACT-M scores were placed into the variable MATH along with the SAT-M scores.

Design

This study was exploratory and correlational, seeking to develop and validate a cognitive model for college mathematics placement and also to compare its effectiveness against other theoretical and empirical placement models. These objectives suggested the following design and hypotheses.

Questions I and II Design

I. Can a cognitive model for college mathematics placement be developed from Skemp's theory of learning and Wilson's model of mathematics achievement?

II. Which predictor variables are logically consistent with such a model for college mathematics placement?

The first two research questions raised in Chapter I were answered by the logical analyses in Chapter II in the sections entitled The Cognitive Model for College Mathematics Placement and Placement Variables, respectively. The following pages describe the design used to answer the last two research questions.

Question III Design

III. Does the Cognitive Model for College Mathematics Placement, using the predictor variables identified in research Question II, produce a valid placement procedure?

The design to answer the third research question involved investigating the reliability and validity of the cognitive placement subscales. The reliability of the subscales was analyzed using measures of internal consistency. The study of the validity of the cognitive subscales was accomplished by investigating their construct, content, and predictive validity. Cross-validations of the prediction equations were also performed. The details of the design follow in the section entitled Question III Procedures.

Question IV Design

IV. Does the Cognitive Model for College Mathematics Placement produce more effective placement than either the Willingham model, the SOP model, or the Computerized Placement model currently being used by DFMS?

The design to answer the fourth research question involved two methods for comparing the efficiency of different placement procedures: (a) comparing the capability of the variables, logically consistent with each procedure, to predict final differential calculus course grade or course assignments, using a generalized F-test, and (b) comparing the number of correct and incorrect placements, actual or hypothetical, produced by each model.

The numbers of correct and incorrect placements for each pair of competing placement models were analyzed using hit-and-miss tables after a suitable hypothetical placement was performed. These hypothetical placements were performed with a cutoff value of the expected final calculus grade derived from data displayed in "Kelley Tables". Kelley Tables present data to support selecting cutoff values in accordance with the guidelines and procedures described by Appenzellar and Kelley (1983).

Procedures

This section contains the details of the procedures used to investigate research Questions III and IV. The research hypotheses in the design section are elaborated and described in operational, statistical forms. No procedures are described here for Questions I and II, as they were already answered by the logical analyses in Chapter II in the sections entitled The Cognitive Model for College Mathematics Placement and Placement Variables, respectively.

Question III Procedures

Content Validity of Cognitive Subscales

The content validities of the placement tests were investigated by having the current DFMS precalculus course director and two precalculus instructors respond to a content validity questionnaire (see Appendix B). The questionnaire contained five-choice Likert scale items which required the judges to classify each of the placement test items as one that did not test a content topic in the syllabus or was either a poor, adequate, good, or an excellent test item of a content topic in the syllabus. These responses were coded as 1 through 5 respectively. The subtest was considered valid if the total average coded value of all items was at least 2.5 and the average percentage of agreement among the judges for all items was at least 67%.

Construct Validity of Cognitive Subscales

A questionnaire, Appendix B, was developed in which each placement test item was classified according to Wilson's (1971) taxonomy. Seven judges were used; two were mathematics education faculty members at a leading university in Texas, two were faculty members at central Texas high schools who had earned doctorate degrees in mathematics education, and three were precalculus instructors in the Department of Mathematical Sciences at USAFA. The judgments of the mathematics education specialists were given more weight in the process of combining the judgments to categorize the items. Thus, an item was categorized as an analysis item if two out of the four mathematics education specialists judged it to be an analysis item, or if any group of judges unanimously categorized it as an analysis item. Two subscales for each of the algebra and trigonometry placement tests were then established such that one subscale contained only analysis items and the other subscale contained only the non-analysis (computation, comprehension, and application) items.

The subscales were used to form two target vectors. One vector represented the lower levels of cognitive behavior specified by the combined classification of each item on a placement test; the value of each element of this vector was 1 if the item was not an analysis item and 0 if the item was an analysis item. The other vector was the complement of the previous vector in that the elements of the vector had a value of 0 if the item was not an analysis item and the value of 1 if the item was an analysis item. These vectors composed the target matrix, H_{nx2}.

Confirmatory factor analysis was then used to assess how well the data fit the theoretical factor pattern. Each of he different placement subscales' item response data from cadets in the study was each factor analyzed using the principal components factor analysis procedure of the SAS (1985) statistics package. The resulting two factor matrix, F_{nx2}, was obliquely rotated using a Procrustes rotation to target technique. The resulting matrix product of the Procrustes transformation matrix, T_{2x2}, and F_{nx2} produced the least squares fit, $H^*_{nx2} = F_{nx2}\ T_{2x2}$ with H_{nx2}. A correlation coefficient statistically different from zero was interpreted as evidence of the construct validity of the subscales. The hypothesis tested was:

Statistical hypotheses H3.1: The Pearson product-moment correlation coefficients between H^*_{nx2} and H_{nx2} are zero.

Reliability of Cognitive Subscales

The reliability of the placement subscales was investigated with KR-20 reliability coefficients computed using the reliability procedure in SPSS-X (1985) . The reliability coefficients for associated examinations were compared across the Classes of 1989, 1990, and 1991 test forms using a test described by Snedecor and Cochran (1969) and attributed to Fisher (1921). Fisher suggested that two correlations can be compared from computing the difference between their associated z^* values. Thus, in this research the pairwise differences of z^* values were analyzed for statistical significance. Bonferroni's method for controlling Type I error was used because of the number of comparisons performed. Thus, the following hypotheses were tested:

Statistical hypothesis H3.2.1: The reliability coefficients of the algebra analysis subscales are not pairwise different for the Classes of 1989, 1990, and 1991.

Statistical hypothesis H3.2.2: The reliability coefficients of the algebra non-analysis subscales are not pairwise different for the Classes of 1989, 1990, and 1991.

Statistical hypothesis H3.2.3: The reliability coefficients of the trigonometry analysis subscales are not pairwise different for the Classes of 1989, 1990, and 1991.

Statistical hypothesis H3.2.4: The reliability coefficients of the trigonometry non-analysis subscales are not pairwise different for the Classes of 1989, 1990, and 1991.

Predictive Validity of Cognitive Variables

The predictive validity of the cognitive variables identified in Question II was investigated in regards to predicting the final calculus grades. This was accomplished by regressing the final calculus grades (GRADE) on the algebra analysis (ALGA) and non-analysis (ALGNA) subscores, the trigonometry analysis (TRIGA) and non-analysis (TRIGNA) subscores, and the equated SAT-M -- ACT-M (MATH) scores. An *F*-test and multiple *t*-tests were used to identify statistically significant predictors. Hence, the following two hypotheses were tested:

Statistical hypothesis H3.3.1: Using the regression equation:

$$GRADE = \beta_0 + \beta_1 ALGA + \beta_2 ALGNA + \beta_3 TRIGA + \beta_4 TRIGNA + \beta_5 MATH + \epsilon \qquad (2)$$

the parameters $\beta_1, \beta_2, \beta_3, \beta_4,$ and β_5 are all simultaneously equal to zero.

Statistical hypothesis H3.3.2: If H3.3.1 is rejected, then some of the parameters $\beta_1, \beta_2, \beta_3, \beta_4,$ and β_5 in Equation 2 are equal to zero.

Cross-Validation of Cognitive model.

The prediction equations for the Classes of 1989, 1990, and 1991 were also cross-validated. The cross-validations were performed by evaluating the shrinkage of the correlation coefficients from the prediction equations for the Classes of 1989, 1990, and 1991. Specifically, correlation coefficients were obtained by the prediction equation obtained from the Class of 1989 data to predict the grades obtained by the Classes of 1990 and 1991, as well as by using the 1990 prediction equation to predict the 1991 grades. The cross-validation shrinkage values were computed using the following formula: shrinkage = $R_{YY'} - R_{XY}$, where $R_{YY'}$ = the maximum possible multiple correlation coefficient between the later year's predicted and observed GRADE values and R_{XY} = product-moment correlation coefficient between the observed values of GRADE for the later year and the predicted values of GRADE using a previous year's prediction model. Shrinkage values less than 0.10 were interpreted as indicating relatively stable accuracies of prediction.

Question IV Procedures

Two methods for comparing placement procedures were used, (a) comparing the capability of the variables, logically consistent with each placement model, to predict final differential calculus course grade or course assignments; and (b) comparing the number of correct and incorrect placements, actual or hypothetical, produced by each model.

The first method used the residual sum of squares (RSS_f) and the degrees of freedom (DF_f) from Equation 2, and the residual sum of squares (RSS_i) and the degrees of freedom (DF_i) from an alternative placement model to calculate an observed F^* value. The F^* value was then compared to the $F(\alpha = 0.05, DF_i - DF_f, DF_f)$ tabled value to test for statistical significance of an improved accuracy of prediction by Equation 2.

The second method used hit-and-miss tables to compare the models. Table 4 was used to define hits and misses on the basis of the relationship between the actual and hypothetical placements and the implications with regards to the correctness of the conjectured placements. Recall that the short sequence refers to placement directly into differential calculus and the long sequence refers to the precalculus-calculus course sequence. A satisfactory grade (S) is one that is at least a C- (grade $\geq$ 2.00); an unsatisfactory grade (U) is one that is less than a C- (grade $<$ 2.00). The number of hits for the

Cognitive and Willingham models were found using the guidelines established in Table 4, that is the sum of the number of cadets in the categories denoted as correct. The number of hits for the empirical models were found by summing the number of cadets who were placed into the long sequence and were unsuccessful and the number of cadets who were placed into the short sequence and were successful.

The conditions where the correctness of the hypothetical placement is unknown occur because it was impossible to determine if students in this hypothetical placement category would have earned satisfactory or unsatisfactory grades if placed directly into the other sequence since the students were not actually placed in the other sequence. For example, suppose that the placement model being investigated suggests that a student should have been placed in the short sequence (calculus), but that the placement model in use that year actually placed the student in the long sequence (precalculus). Further suppose that the student subsequently made a satisfactory grade in calculus. One cannot determine how the student would have performed in calculus had the student been placed directly into that course.

Table 4

General Hit-and-Miss Table Interpretation

Hypothetical Placement	Actual Placement			
	Long Sequence		Short Sequence	
	Performance		Performance	
	U[a]	S[b]	U	S
Long Sequence	Correct[c]	Unknown	Correct	Incorrect
Short Sequence	Incorrect	Unknown	Incorrect	Correct

Note. [a] Unsatisfactory observed grade in calculus.
[b] Satisfactory observed grade in calculus. [c] Correctness of hypothetical placement.

The comparison of the hit-and-miss tables produced from the competing models were performed. The model with the largest percentage of hits was the more efficient model.

Comparing the Cognitive and Willingham Models

Willingham models were constructed for all three classes and validated. Predictive equations regressing GRADE on ALG and TRIG were developed for each class. The general form of these models was:

$$GRADE = \beta_0 + \beta_1 ALG + \beta_2 TRIG + \varepsilon \qquad (3).$$

Validating Willingham models.

The Willingham models were validated using evidence from content and predictive validations and also cross-validated across the three classes. Finally, the internal consistency of the placement tests was analyzed across all three years. The procedures for the content and predictive validation, cross-validation, and reliability were the same that were used for the Cognitive model. Recall that shrinkage values of less than 0.10 were interpreted as indicating relative stability of prediction across the classes. The appropriate statistical hypotheses are listed below.

Statistical hypothesis H4.1.1: The reliability coefficients of the algebra placement tests are not pairwise different for the Classes of 1989, 1990, and 1991.

Statistical hypothesis H4.1.2: The reliability coefficients of the trigonometry placement tests are not pairwise different for the Classes of 1989, 1990, and 1991.

Statistical hypothesis H4.2.1: The parameters β_1 and β_2 of Equation 3 are both simultaneously equal to zero.

Statistical hypothesis H4.2.2: If H4.2.1 is rejected, then some of the parameters β_1 and β_2 in Equation 3 are equal to zero.

<u>Cognitive and Willingham models hit-and-miss tables comparison.</u>

The hit-and-miss tables produced by the Cognitive and Willingham models were constructed after cutoff scores were identified and the classes were hypothetically placed. The cutoff scores and hypothetical placements were accomplished using the procedures of Appenzellar and Kelley (1983).

These methods required six potential cutoff Composite Scores to be developed:

1. The Expected Composite Score for those students with a Final Course Grade of 2.00.
2. The Composite Score for those students with an Expected Final Course Grade of 2.00.
3. The Composite Score for which the percentages of errors of students in each academic performance category (Satisfactory or Unsatisfactory), were most nearly equal.
4. The Composite Score for which the overall percentages of errors were most nearly equal.
5. The Composite Score that would have cut off, or held back, approximately the same number of students as were in the Unsatisfactory performance category.
6. The Composite Score that would have maximized the overall accuracy of placement.

The first guideline score value was obtained from regressing the Composite Scores obtained from regression Equation 2 for the

Cognitive model, or Equation 3 for the Willingham model, on the observed Final Calculus Grades. This model may be written;

$$\text{Composite Score} = \beta_0 + \beta_1\text{GRADE} + \varepsilon \quad (4).$$

The desired guideline value was obtained by substituting GRADE = 2.00 into the regression equation.

The second guideline score value was always 2.00, as it represents the regression of the Final Calculus Grade on the Composite Scores. It is easy to show that the correlation between the Expected Final Calculus Grade and the Composite Score is 1.0. Thus, this guideline produced an Expected Final Calculus Grade of 2.00 in all cases.

The third and fourth guideline score values were obtained from tables of Composite Scores with various possible cutting Composite Scores and corresponding placement accuracies (see Table D.4 in Appendix D). The fifth and sixth guideline score values were obtained from tables of Composite Scores by *observed Final Calculus Grades* aggregated into Satisfactory and Unsatisfactory categories (see Table D.3 in Appendix D).

The most appropriate cutoff Composite Score was selected, after examining the six guideline values, based on distributional and reasonability factors. The cutoff Composite Scores were then used to determine the hypothetical placement of the students into the long and short sequences, based on the cadets' Composite Scores in relation to the cutoff Composite Scores.

Cognitive and Willingham models generalized F-tests comparison.

The Cognitive and the Willingham models were compared to determine which explained more of the variance in the data. The Cognitive regression Equation 2 was compared to the Willingham regression Equation 3, using a generalized F-test. For this test RSS_1

and DF_l were the residual sum of squares and the degrees of freedom from the Willingham model and RSS_f and DF_f were the residual sum of squares and the degrees of freedom from the Cognitive model. The calculated F^* value was compared to the $F(\alpha = 0.05, DF_l - DF_f, DF_f)$ tabled value to test for significance. An observed F^* value greater than the tabled F value was interpreted as indicating that the Cognitive model explained more of the variance in the data than did the Willingham model. Thus, the following hypothesis was tested:

Statistical hypothesis H4.3: In Equation 2, $\beta_1 = \beta_2, \beta_3 = \beta_4$, and $\beta_5 = 0$.

Cognitive and SOP Models Comparison

Since the SOP placement model was the actual placement procedure for the Classes of 1989 and 1990, the hit-and-miss tables for the SOP model were obtained directly from the data. These tables were compared to the Cognitive model hit-and-miss tables developed for the Cognitive -- Willingham models comparison procedure.

Cognitive and Computerized Models Comparison

DFMS personnel reported that there were significant numbers of cadets in the Class of 1991 who were placed into sequences other than that prescribed by the Computerized model. This researcher determined these "hand-placed" cadets to be a threat to the validity of any placement model comparison and so these cadets were removed from the analysis. The procedure for removing the hand-placed cadets involved determining the original Computerized Placement Model placement, comparing that to the observed placement, and removing those cadets for whom the original and observed placements disagreed. The original placement was obtained from DFMS, and the 1991 data set was appropriately reduced.

This reduced Class of 1991 was then used to compare the Cognitive and Computerized models. The models were compared using both a generalized F-test and hit-and-miss tables.

Cognitive and Computerized models hit-and-miss table comparison.

The Cognitive model was compared with the Computerized model using a hit-and-miss table The hit-and-miss table for the Computerized model was directly obtained from the reduced 1991 data set. The hit-and-miss table for the Cognitive model was obtained by applying the basic Cognitive regression Equation 2 to the reduced 1991 data set. A cutoff Expected Final Calculus Grade was obtained using the methods of Appenzellar and Kelley (1983), as before. The procedures for performing the hypothetical placement, preparing the hit-and-miss tables, and comparing the hit-and-miss tables were consistent with the previous procedures.

Cognitive and Computerized models generalized F-tests comparison.

Under the assumption of multivariate normality of the prediction variables, multiple linear regression of binary data produces a discriminant function equivalent to linear discriminant function analysis (Morris & Rolph, 1981). It was determined that the Cognitive model could be compared to the Computerized model by using a generalized F-test. The Computerized model for the non-USAFA Preparatory School graduates was represented by the regression equation:

$$PASS = \beta_0 + \beta_1 CALC2 + \beta_2 ALG + \beta_3 ACI + \beta_4 CALC1 + \beta_5 TRIG + \beta_6 CALC3 + \varepsilon \quad (5).$$

No other equation was used for the other groups in the class of 1991 because there was an enforced DFMS policy that did not allow USAFA

Preparatory School graduates to be placed into precalculus, and another DFMS policy that did not allow non-USAFA Preparatory School graduates who scored less than 45% on the algebra placement test, to be placed in any course but precalculus.

The Cognitive model counterpart to Equation 5 was developed by replacing the appropriate non-cognitive variables with cognitive variables. This augmented Computerized regression equation was:

$$PASS = \beta_0 + \beta_1 CALC2 + \beta_2 ALGA + \beta_3 ALGNA + \beta_4 ACI + \beta_5 CALC1 + \beta_6 TRIGA + \beta_7 TRIGNA + \beta_8 MATH + \beta_9 CALC3 + \varepsilon \quad (6),$$

where PASS was the variable formed by classifying cadets as successful (i.e., GRADE $\geq$ 2.00) or unsuccessful (i.e., GRADE < 2.00).

The Augmented Computerized regression Equation 6 was compared to the Computerized regression Equation 5 using a generalized F-test. For this test RSS_i and DF_i were the residual sum of squares and the degrees of freedom from the Computerized model and RSS_f and DF_f were the residual sum of squares and the degrees of freedom from the Augmented Computerized model. The calculated F^* value was compared to the $F(\alpha = 0.05, DF_i - DF_f, DF_f)$ tabled value to test for significance. An observed F value greater than the tabled F value was interpreted as indicating that the Augmented Computerized model explained more of the variance in the data than the Computerized model due to the addition of the cognitive variables.

Statistical hypothesis H4.2.1: In Equation 6, $\beta_2 = \beta_3 = \beta_6 = \beta_7 = \beta_8 = 0$.

Statistical hypothesis H4.2.2: If H4.2.1 is rejected, then some of β_2, β_3, β_6, β_7, and β_8 in Equation 6 are equal to zero.

Statistical hypothesis H4.2.3: In Equation 6 $\beta_2 = \beta_3$, $\beta_6 = \beta_7$, and $\beta_8 = 0$.

CHAPTER IV

RESULTS

This chapter describes the results of executing the procedures listed in the Methodology Chapter. The results presented in this chapter focus on supplying relevant information about research Questions III and IV, as research Questions I and II were previously answered in the sections entitled The Cognitive Model for College Mathematics Placement and Placement Variables, respectively.

Descriptive Statistics

Basic descriptive statistics for all the relevant variables in the study are presented below. The means, standard deviations, and intercorrelations for all the dependent and independent variables are reported for each of the classes in the study broken down by placement sequence.

Table 5.1.0 through Table 5.3.5 emphasize the similarities between the three classes of cadets in terms of most of the variables in the study. However, an apparent difference between the classes seemed to exist for the means of the final calculus grades (GRADE). The large samples normal approximation to the $\underline{F}$-test,

$$\underline{u} = \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2(n_2-1)} + \frac{s_2^2}{2(n_1-1)}}},$$

was first used to test if the class variances were the same, since all classes had at least 582 observations. This test showed that each of the variances was significantly different at the $\alpha = 0.05$ level. This result meant that the differences between the means had to be tested using "Behrens-Fisher" methods. In this case, Burr (1974) suggested

Table 5.1.0
1989 Means

Variable	n	Mean	SD
ACI[a]	582	3068.40	229.41
ACT-M	323	28.64	3.04
ALG[b]	582	22.50	5.79
ALGA	582	3.70	1.28
ALGNA	582	18.81	5.04
GRADE	582	2.91	0.77
MATH[c]	582	619.92	63.90
PAR	582	620.74	84.40
SAT-M	259	639.46	49.84
TRIG[d]	582	12.41	3.49
TRIGA	582	1.06	0.70
TRIGNA	582	11.35	3.17

Note. [a] ACI is derived in part from ACT-M/SAT-M and PAR. [b] ALG = ALGA + ALGNA. [c] MATH is derived from ACT-M and SAT-M. [d] TRIG = TRIGA + TRIGNA.

Table 5.1.1
1989 Intercorrelations

	ACI	ACT-M	ALG	ALGA	ALGNA	GRADE	MATH	PAR	SAT-M	TRIG	TRIGA	TRIGNA
ACI	1.00	.52	.28	.18	.27	.18	.47	.79	.44	.20	.09	.20
ACT-M		1.00	.47	.34	.45	.18	1.00	-.02	--	.24	.24	.21
ALG			1.00	.66	.98	.14	.44	.04	.44	.48	.31	.46
ALGA				1.00	.51	.06	.32	-.02	.33	.31	.27	.29
ALGNA					1.00	.14	.43	.05	.42	.48	.28	.46
GRADE						1.00	.19	.10	.18	.18	.13	.17
MATH							1.00	-.04	1.00	.25	.24	.22
PAR								1.00	-.06	.08	-.05	.10
SAT-M									1.00	.24	.18	.23
TRIG										1.00	.54	.98
TRIGA											1.00	.38
TRIGNA												1.00

Table 5.1.2
1989 Precalculus-Calculus Sequence Means

Variable	n	Mean	SD
ACI[a]	121	2992.79	223.46
ACT-M	70	27.06	2.10
ALG[b]	121	15.70	4.04
ALGA	121	2.70	1.14
ALGNA	121	13.00	3.56
GRADE	121	3.10	0.84
MATH[c]	121	590.74	56.22
PAR	121	614.14	94.58
SAT-M	51	623.33	51.52
TRIG[d]	121	10.09	3.44
TRIGA	121	0.84	0.69
TRIGNA	121	9.26	3.11

Note. [a] ACI is derived in part from ACT-M/SAT-M and PAR. [b] ALG = ALGA + ALGNA. [c] MATH is derived from ACT-M and SAT-M. [d] TRIG = TRIGA + TRIGNA.

Table 5.1.3
1989 Precalculus-Calculus Sequence Intercorrelations

	ACI	ACT-M	ALG	ALGA	ALGNA	GRADE	MATH	PAR	SAT-M	TRIG	TRIGA	TRIGNA
ACI	1.00	.07	.15	.04	.16	.05	.32	.83	.55	.16	-.15	.21
ACT-M		1.00	.27	.33	.19	.16	1.00	-.37	--	-.02	.02	-.03
ALG			1.00	.54	.96	.19	.30	-.03	.44	.40	.24	.39
ALGA				1.00	.30	.15	.26	-.14	.25	.19	.19	.17
ALGNA					1.00	.16	.25	.01	.43	.39	.22	.39
GRADE						1.00	.12	.08	-.11	.05	-.04	.06
MATH							1.00	-.16	1.00	.11	.10	.10
PAR								1.00	.03	10	-.22	.16
SAT-M									1.00	.15	.14	.14
TRIG										1.00	.56	.98
TRIGA											1.00	.40
TRIGNA												1.00

Table 5.1.4

1989 Calculus Sequence Means

Variable	n	Mean	SD
ACI[a]	461	3088.24	227.05
ACT-M	253	29.07	3.12
ALG[b]	461	24.29	4.77
ALGA	461	3.96	1.18
ALGNA	461	20.33	4.19
GRADE	461	2.86	0.74
MATH[c]	461	627.58	63.64
PAR	461	622.47	81.54
SAT-M	208	643.41	48.74
TRIG[d]	461	13.02	3.25
TRIGA	461	1.12	0.69
TRIGNA	461	11.90	2.95

Note. [a] ACI is derived in part from ACT-M/SAT-M and PAR. [b] ALG = ALGA + ALGNA. [c] MATH is derived from ACT-M and SAT-M. [d] TRIG = TRIGA + TRIGNA.

Table 5.1.5

1989 Calculus Sequence Intercorrelations

	ACI	ACT-M	ALG	ALGA	ALGNA	GRADE	MATH	PAR	SAT-M	TRIG	TRIGA	TRIGNA
ACI	1.00	.56	.24	.15	.23	.25	.48	.78	.40	.16	.12	.14
ACT-M		1.00	.41	.26	.40	.24	1.00	.03	--	.19	.25	.15
ALG			1.00	.59	.97	.29	.40	.03	.44	.36	.27	.34
ALGA				1.00	.39	.11	.26	-.01	.31	.21	.23	.17
ALGNA					1.00	.30	.39	.04	.41	.35	.24	.33
GRADE						1.00	.26	.12	.28	.30	.21	.28
MATH							1.00	-.02	1.00	.20	.23	.17
PAR								1.00	-.09	.07	0.00	.07
SAT-M									1.00	.22	.16	.21
TRIG										1.00	.52	.98
TRIGA											1.00	.33
TRIGNA												1.00

Table 5.2.0
1990 Means

Variable	n	Mean	SD
ACI[a]	692	3055.97	217.78
ACT-M	374	28.04	2.73
ALG[b]	692	22.90	6.28
ALGA	692	3.72	1.32
ALGNA	692	19.19	5.45
GRADE	692	2.52	0.96
MATH[c]	692	612.67	62.86
PAR	692	622.63	79.72
SAT-M	318	638.71	52.72
TRIG[d]	692	12.28	3.63
TRIGA	692	1.03	0.68
TRIGNA	692	11.24	3.33

Note. [a] ACI is derived in part from ACT-M/SAT-M and PAR. [b] ALG = ALGA + ALGNA. [c] MATH is derived from ACT-M and SAT-M. [d] TRIG = TRIGA + TRIGNA.

Table 5.2.1
1990 Intercorrelations

	ACI	ACT-M	ALG	ALGA	ALGNA	GRADE	MATH	PAR	SAT-M	TRIG	TRIGA	TRIGNA
ACI	1.00	.45	.24	.14	.25	.25	.42	.78	.45	.10	.04	.10
ACT-M		1.00	.35	.21	.35	.21	1.00	-.09	--	.14	.14	.13
ALG			1.00	.70	.99	.34	.39	.05	.48	.52	.26	.51
ALGA				1.00	.56	.21	.23	.03	.29	.42	.20	.42
ALGNA					1.00	.34	.39	.05	.49	.49	.25	.49
GRADE						1.00	.23	.14	.30	.25	.13	.25
MATH							1.00	-.10	1.00	.14	.14	.12
PAR								1.00	-.06	.03	-.06	.05
SAT-M									1.00	.22	.17	.20
TRIG										1.00	.51	.98
TRIGA											1.00	.36
TRIGNA												1.00

Table 5.2.2

1990 Precalculus-Calculus Sequence Means

Variable	n	Mean	SD
ACI[a]	123	2955.50	202.40
ACT-M	62	27.07	2.40
ALG[b]	123	14.57	3.78
ALGA	123	2.42	1.07
ALGNA	123	12.15	3.44
GRADE	123	2.32	1.02
MATH[c]	123	584.50	50.51
PAR	123	606.90	74.23
SAT-M	61	600.00	40.21
TRIG[d]	123	8.64	3.10
TRIGA	123	0.71	0.66
TRIGNA	123	7.94	2.86

Note. [a] ACI is derived in part from ACT-M/SAT-M and PAR. [b] ALG = ALGA + ALGNA. [c] MATH is derived from ACT-M and SAT-M. [d] TRIG = TRIGA + TRIGNA.

Table 5.2.3

1990 Precalculus-Calculus Sequence Intercorrelations

	ACI	ACT-M	ALG	ALGA	ALGNA	GRADE	MATH	PAR	SAT-M	TRIG	TRIGA	TRIGNA
ACI	1.00	.56	.11	.03	.12	.34	.48	.78	.55	.01	-.13	.04
ACT-M		1.00	.18	-.13	.23	.30	1.00	-.05	--	.03	.15	0.00
ALG			1.00	.44	.96	.14	.16	-.02	.26	.19	.12	.18
ALGA				1.00	.18	.10	-.03	.06	.08	.19	.09	.19
ALGNA					1.00	.12	.18	-.04	.27	.15	.10	.14
GRADE						1.00	.19	.22	.21	.02	-.09	.04
MATH							1.00	-.04	1.00	-.03	.04	-.04
PAR								1.00	.07	-.01	-.16	.04
SAT-M									1.00	-.05	-.17	-.02
TRIG										1.00	45	.98
TRIGA											1.00	.26
TRIGNA												1.00

Table 5.2.4

1990 Calculus Sequence Means

Variable	n	Mean	SD
ACI[a]	569	3078.00	215.00
ACT-M	312	28.23	2.75
ALG[b]	569	24.71	5.16
ALGA	569	4.00	1.20
ALGNA	569	20.71	4.53
GRADE	569	2.57	0.94
MATH[c]	569	618.80	63.64
PAR	569	626.00	80.52
SAT-M	257	647.90	51.19
TRIG[d]	569	13.06	3.23
TRIGA	569	1.10	0.66
TRIGNA	569	11.96	2.97

Note. [a] ACI is derived in part from ACT-M/SAT-M and PAR. [b] ALG = ALGA + ALGNA. [c] MATH is derived from ACT-M and SAT-M. [d] TRIG = TRIGA + TRIGNA.

Table 5.2.5

1990 Calculus Sequence Intercorrelations

	ACI	ACT-M	ALG	ALGA	ALGNA	GRADE	MATH	PAR	SAT-M	TRIG	TRIGA	TRIGNA
ACI	1.00	.42	.15	.05	.16	.21	.38	.78	.37	0.00	.01	-.01
ACT-M		1.00	.34	.19	.33	.19	1.00	-.11	--	.09	.10	.07
ALG			1.00	.61	.98	.40	.36	-.01	.36	.35	.16	.34
ALGA				1.00	.44	.21	.19	-.02	.14	.28	.11	.28
ALGNA					1.00	.40	.36	-.01	.38	.32	.15	.32
GRADE						1.00	.23	.11	.25	.28	.16	.27
MATH							1.00	-.14	1.00	.06	.11	.04
PAR								1.00	-.12	-.01	-.07	0.00
SAT-M									1.00	.07	.15	.04
TRIG										1.00	.48	.98
TRIGA											1.00	.30
TRIGNA												1.00

Table 5.3.0

1991 Means

Variable	n	Mean	SD
ACI[a]	742	3065.65	228.32
ACT-M	384	28.02	2.86
ALG[b]	742	22.85	6.30
ALGA	742	3.80	1.34
ALGNA	742	19.05	5.49
GRADE	742	2.47	1.07
MATH[c]	742	613.37	62.26
PAR	742	625.18	81.61
SAT-M	358	638.07	47.37
TRIG[d]	742	12.64	3.60
TRIGA	742	1.04	0.70
TRIGNA	742	11.60	3.26

Note. [a] ACI is derived in part from ACT-M/SAT-M and PAR. [b] ALG = ALGA + ALGNA. [c] MATH is derived from ACT-M and SAT-M. [d] TRIG = TRIGA + TRIGNA.

Table 5.3.1

1991 Intercorrelations

	ACI	ACT-M	ALG	ALGA	ALGNA	GRADE	MATH	PAR	SAT-M	TRIG	TRIGA	TRIGNA
ACI	1.00	.51	.40	.28	.39	.38	.47	.84	.53	.31	.17	.31
ACT-M		1.00	.52	.41	.50	.34	1.00	.06	--	.33	.33	.30
ALG			1.00	.67	.98	.44	.44	.25	.42	.53	.35	.51
ALGA				1.00	.53	.37	.33	.16	.28	.37	.26	.35
ALGNA					1.00	.42	.42	.25	.41	.51	.34	.49
GRADE						1.00	.27	.31	.28	.40	.29	.37
MATH							1.00	.04	1.00	.27	.26	.24
PAR								1.00	.11	.21	.06	.22
SAT-M									1.00	.31	.24	.29
TRIG										1.00	.56	.98
TRIGA											1.00	.41
TRIGNA												1.00

Table 5.3.2

1991 Precalculus-Calculus Sequence Means

Variable	n	Mean	SD
ACI[a]	185	2925.00	209.74
ACT-M	94	26.38	1.84
ALG[b]	185	15.72	3.72
ALGA	185	2.84	1.24
ALGNA	185	12.89	3.34
GRADE	185	1.87	1.16
MATH[c]	185	579.46	46.81
PAR	185	591.42	81.85
SAT-M	91	606.59	36.73
TRIG[d]	185	9.42	3.39
TRIGA	185	0.69	0.68
TRIGNA	185	8.72	3.12

Note. [a] ACI is derived in part from ACT-M/SAT-M and PAR. [b] ALG = ALGA + ALGNA. [c] MATH is derived from ACT-M and SAT-M. [d] TRIG = TRIGA + TRIGNA.

Table 5.3.3

1991 Precalculus-Calculus Sequence Intercorrelations

	ACI	ACT-M	ALG	ALGA	ALGNA	GRADE	MATH	PAR	SAT-M	TRIG	TRIGA	TRIGNA
ACI	1.00	.37	.01	.10	-.02	.29	.34	.86	.42	.02	-.06	.03
ACT-M		1.00	.12	.13	.08	.25	.99	-.01	--	.03	.20	-.02
ALG			1.00	.46	.94	.14	.12	-.02	.06	.06	-.05	.07
ALGA				1.00	.14	.19	.16	.04	.12	.13	.01	.14
ALGNA					1.00	.08	.07	-.03	.03	.02	-.06	.03
GRADE						1.00	.12	.23	.14	.23	.20	.20
MATH							1.00	-.04	1.00	-.04	.09	-.06
PAR								1.00	-.03	.03	-.08	.05
SAT-M									1.00	-.07	-.06	-.06
TRIG										1.00	.49	.98
TRIGA											1.00	.31
TRIGNA												1.00

Table 5.3.4
1991 Calculus Sequence Means

Variable	n	Mean	SD
ACI[a]	557	3112.37	214.82
ACT-M	290	28.55	2.93
ALG[b]	557	25.22	5.08
ALGA	557	4.12	1.22
ALGNA	557	21.10	4.43
GRADE	557	2.66	0.96
MATH[c]	557	624.63	62.70
PAR	557	636.39	78.45
SAT-M	267	648.80	45.83
TRIG[d]	557	13.72	2.98
TRIGA	557	1.16	0.67
TRIGNA	557	12.56	2.69

Note. [a] ACI is derived in part from ACT-M/SAT-M and PAR. [b] ALG = ALGA + ALGNA. [c] MATH is derived from ACT-M and SAT-M. [d] TRIG = TRIGA + TRIGNA.

Table 5.3.5
1991 Calculus Sequence Intercorrelations

	ACI	ACT-M	ALG	ALGA	ALGNA	GRADE	MATH	PAR	SAT-M	TRIG	TRIGA	TRIGNA
ACI	1.00	.47	.29	.18	.29	.30	.43	.82	.47	.21	.12	.20
ACT-M		1.00	.48	.35	.46	.29	1.00	-.03	--	.25	.27	.21
ALG			1.00	.62	.98	.39	.36	.17	.28	.36	.30	.33
ALGA				1.00	.44	.31	.25	.07	.17	.23	.21	.20
ALGNA					1.00	.36	.34	.17	.27	.35	.28	.32
GRADE						1.00	.21	.26	.19	.31	.23	.29
MATH							1.00	-.04	1.00	.18	.21	.14
PAR								1.00	.04	.13	.02	.14
SAT-M									1.00	.20	.22	.16
TRIG										1.00	.51	.98
TRIGA											1.00	.32
TRIGNA												1.00

using the Aspin-Welch procedure for comparing means. With this procedure the pairwise differences between the mean GRADE values, listed in Table 6, were compared using the test statistic:

$$\underline{t} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

The data in Table 6 indicate that, on the average, the final course grades earned by the Class of 1989 were significantly higher than those earned by both the Class of 1990 and the Class of 1991, even after using the Bonferroni procedure for controlling inflated Type I error.

Table 6

Pairwise Comparison of Mean GRADE by Class

Comparison	Difference	df[a]	t
1989-1990	0.39	1269	7.92**
1989-1991	0.44	1312	8.73**
1990-1991	0.05	1430	1.08

Note. [a] Degrees of freedom. ** Significant at the $\alpha = 0.01$ level.

These significant differences could have resulted from the differing initial capabilities of the classes, but this does not seem reasonable given the similarities of the other achievement measures. Another possible reason could be that DFMS reacted to grade inflation with the Class of 1989 and consciously corrected the problem by controlling the grade distribution of the final calculus grades. This seems plausible since only 11 out of 582 (1.9%) cadets in the long and short sequences earned Ds and Fs in the Class of 1989, while, for the Class of 1991, the unsuccessful rate increased to 103 out of 742 (13.9%). The distributions for all grades are displayed for each class in Table 7.

Table 7

Final Course Grades in Calculus by Class

Final Course Grades in Calculus[a]

Year	F	D	C	B	A	Total
1989	4(0.7)	7(1.2)	156(26.8)	287(49.3)	128(22.0)	582(100)
1990	25(3.6)	48(6.9)	269(38.9)	240(34.7)	110(15.9)	692(100)
1991	50(6.7)	53(7.1)	271(36.5)	238(32.1)	130(17.5)	742(100)

Note. [a] The number of cadets who earned the grade (percentage of class in the study).

Table 7, in conjunction with Table 3, contains evidence of another possible DFMS policy change. The number of cadets placed into either the long or short sequence increased by 90 cadets from the Class of 1989 to the Class of 1990 and 50 cadets from the Class of 1990 to the Class of 1991. This means that the enrollment for the Class of 1989, 42.3% of the total entering freshman class, increased to 55.0% of the total entering freshman Class of 1991. This translates into an increase of 12.7%, with about one extra cadet placed into the long sequence for every two extra placed into the short sequence. Thus, over the three class period, there were fewer advanced placed freshmen and cadets, who were likely to be placed out of the short sequence in the Class of 1989, but were more likely to placed into the short sequence in the Class of 1991. These cadets were also less likely to receive as high a final grade in calculus as those in the Class of 1989. These changes in course placement and final calculus distributions demonstrate a changing environment commensurate with implementing new policies.

While the methodology chapter implied that the long and short sequences were stable in terms of the placement procedures used to enroll cadets, the capability of cadets enrolled, syllabi, examinations, grading of the examinations, and assigning grades, it appears that changes did indeed take place. Apparently, many of the cadets in the Classes of 1990 and 1991 would have been placed differently had they been in the Class of 1989. In addition, there may have been an effort to counteract grade inflation in the calculus course. These factors will necessarily impact the results of the investigation of research Question III and Question IV.

Question III Results

> III. Does the Cognitive Model for College Mathematics Placement, using the predictor variables identified in research Question II, produce a valid placement procedure?

Research Question III was concerned about the validity of the Cognitive Model for College Mathematics Placement in conjunction with the placement instruments used by DFMS for the Classes of 1989, 1990, and 1991. The current section reports the results of investigating the construct, content, and predictive validities of the Cognitive model for each class.

<u>Content Validity of Cognitive Subscales</u>

The content validity of the placement subscales was investigated using judges' responses on the Content Validity Questionnaire (Appendix B). The judges responded to a single Likert scale item which required each judge to classify each placement subscale item as one that did not test a content topic in the syllabus, or was either a poor, adequate, good, or an excellent test item of a content topic in the syllabus. These responses were coded as _1_ through _5_ respectively. The subscales were considered content valid if the average coded value was at least 2.5 and the average percentage of agreement among the judges was at least 67%. The results of the content validity study are presented in Table 8.

Table 8
Cognitive Model Content Validity Agreement

Test	No. of Items	Average % Agreement	Modal Judgment	Average Judgment
ALG	40	65.0	4	4.0
ALGA	6	33.3	4	3.7
ALGNA	34	70.6	4	4.0
TRIG	20	75.0	4	4.1
TRIGA	2	83.3	NM [a]	4.2
TRIGNA	18	74.1	4	4.0

Note. [a] No mode.

As shown in Table 8, only the algebra analysis subscale did not meet the content validity criterion. Even though the average coded score of 3.7 exceeded the criterion of 2.5, the average percentage of agreement, 33.3% was smaller than the required 67%. Appendix B shows that three out of the six items (50%) composing the algebra analysis subscale had no agreement among the judges concerning the adequacy of the items. The three items which had 0% agreement had score patterns of coded judgments of 3, 4, and 5 indicating that while a true disagreement existed among the judges, the disagreement was not about the whether the items were not representative of the content of the course nor that the items were poor test items. All of the judges felt that these three items were at least adequate test items. Without these three average coded scores the average percent agreement was 67%. Thus, the algebra analysis and non-analysis subscales were composed of items which were all judged to be at least adequate test items of the content of the precalculus course.

The trigonometry cognitive subscales received judgments which exceeded all of the content validity criteria. This test can be characterized as having good test items which reflect the trigonometry content of the precalculus course.

Construct Validity of Cognitive Subscales

> H3.1: The Pearson product-moment correlation coefficients between the hypothesized target factor loadings and the observed two-factor factor loadings, after rotation, are zero.

The results of testing H3.1 are reported in this section. This hypothesis seeks to investigate the construct validity of the DFMS placement examinations interpreted by the Cognitive Model for College Mathematics Placement.

The construct validity was investigated by classifying the items of the algebra and trigonometry placement examinations as either analysis or non-analysis items and then performing confirmatory factor analysis of cadets' individual responses, separated into analysis and non-analysis subscales. The confirmatory factor analysis was performed using a principal components factor analysis procedure with the two factor solution subsequently obliquely rotated using the Procrustes rotation option of SAS (1985). The subsequent rotated factor pattern was correlated with the target identified by the analysis/non-analysis item pattern of the placement examinations.

The analysis/non-analysis item pattern was determined from combining the judgments of mathematics and mathematics education experts about the algebra and trigonometry placement test items placed on a questionnaire (see Appendix B). The judges were asked to classify the items as either computation, comprehension, application, or analysis items, based on the major behavior classification descriptions of Wilson's (1971) Model of Mathematics Achievement. The judgments were then combined to form analysis and non-analysis item subscales for each content strand. Items were primarily categorized using the judgments of the mathematics education experts; an item was classified as an analysis item if at least 50% of them agreed on the category. The items categorized as analysis items are listed in Table 9 with their associated percentage of agreements. A complete list of all the items and their classifications are compiled in Appendix C.

The items identified in Table 9 defined four subscales. Algebra placement test items numbers 6, 11, 15, 18, 26, and 39 formed the algebra analysis subscale (ALGA), and the remainder of the algebra items formed the algebra non-analysis subscale (ALGNA). The trigonometry placement test items numbers 13 and 19 formed the trigonometry analysis subscale (TRIGA) and the remainder of the trigonometry items formed the trigonometry non-analysis subscale (TRIGNA).

The subscales were used to form two target vectors for each content area for the subsequent factor analysis. One vector, for each content area, represented the lower three main classification levels of cognitive behavior (computation, comprehension, and application), specified by the non-analysis subscale. The values of the elements of this vector were _1_ if the items _were not_ analysis items and _0_ if the items _were_ analysis items. The other vector, for each content area, was the complement of the previous vector. These vectors were used as targets for an oblique Procrustes rotation of the two factor, principal components factor solution of the individual cadet responses for each class by placement test. The resulting factor loadings were then correlated with the original target vectors to assess the similarity with the hypothesized factor pattern. The results of the factor analyses, rotations, and correlations are given in Table 10.

Table 9

Percent Agreement of Analysis Items

Algebra Item No.[a]:	6	11	15	18	26	39
% agreement[b]:	43	57	86	86	43	57

	% Agreement
Total Algebra	
Complete Test	73.7
Analysis Subscale	62.0
Non-analysis subscale	75.7

Trigonometry Item No.[c]:	13	19
% agreement:	57	29

	% Agreement
Total Trigonometry	
Complete Test	66.4
Analysis Subscale	43.0
Non-analysis Subscale	69.0

Note. [a] Algebra items classified as analysis items. [b] Percentage of agreement of behavior level classifications. [c] Trigonometry items classified as analysis items.

Table 10 reveals that the first two factors of the solution space explain only a small amount of variance in the data for each of content areas. In addition, after rotation the reference axes were virtually identical which suggests that the first two factors did not correspond to the hypothesized analysis and non-analysis factors. Thus, it is reasonable that the correlations between the resulting factor patterns and the target factor patterns were not significant at the $\alpha = 0.05$ level. Therefore, there was no evidence to reject hypothesis H3.1, providing little support for the decomposition of the placement examinations into analysis and non-analysis subscales.

Table 10
Summary of Results of Confirmatory Factor Analyses

Test/Class	% Variance Explained[a]	Total Communality	Factors Correlation[b]	Factor Loadings Correlations and t-values[c] Factor 1	Factor 2
ALG 1989	9.8(4.7)	5.80	0.898	-0.06 (-.35)	0.15 (.94)
ALG 1990	11.6(4.1)	6.24	0.956	0.01 (.09)	0.03 (-.16)
ALG 1991	11.5(3.8)	6.13	0.761	0.01 (.06)	0.18 (1.11)
TRIG 1989	16.2(8.5)	4.93	0.889	-0.04 (-.19)	0.13 (.56)
TRIG 1990	18.5(7.6)	5.21	0.869	0.13 (.57)	0.02 (.08)
TRIG 1991	17.3(7.9)	5.04	0.774	-0.06 (-.24)	0.20 (.85)

Note. [a] % variance explained by Factor 1 (% variance explained by Factor 2) eliminating other factors. [b] Correlations between reference axes of the rotated factor space. [c] Correlations between target and observed factor loadings (t-value for test of H_0: $r = 0$).

Thus, the analysis subscales did not seem to be measuring precisely what the experts thought; however, the content validity investigation shows that the trait being measured is connected to the precalculus course. The next concern was to establish the consistency of the measurement which is discussed in the next section.

Reliability of Cognitive Subscales

This section reports the results of investigating the reliability, internal consistency, of the cognitive subscales for adequacy and stability over time. The statistical hypotheses tested in this section were:

Statistical hypothesis H3.2.1: The reliability coefficients of the algebra analysis subscales are not pairwise different for the classes of 1989, 1990, and 1991.

Statistical hypothesis H3.2.2: The reliability coefficients of the algebra non-analysis subscales are not pairwise different for the classes of 1989, 1990, and 1991.

Statistical hypothesis H3.2.3: The reliability coefficients of the trigonometry analysis subscales are not pairwise different for the classes of 1989, 1990, and 1991.

Statistical hypothesis H3.2.4: The reliability coefficients of the trigonometry non-analysis subscales are not pairwise different for the classes of 1989, 1990, and 1991.

The KR-20 Reliability coefficients, estimates of internal consistency, were calculated for the placement examinations for each class using only the scores from the cadets in the study; these values are reported in Table 11. Overall, the reliability values for the analysis subscales of the algebra and trigonometry placement tests were all low, with the algebra analysis items explaining at least 28% of the variance of their subscale total scores, and the trigonometry analysis items explaining at least 17% of the variance of their subscale total scores. The total tests and the non-analysis subscales show evidence of acceptable reliability values.

The reliability coefficients for appropriate examinations were pairwise compared across the classes of 1989, 1990, and 1991 using a test described by Snedecor and Cochran (1969) and attributed to Fisher (1921). This procedure uses the logarithmic transformation of the correlation coefficient r to $z^* = \frac{1}{2}\ln\left(\frac{1 + r}{1 - r}\right)$ such that the standard error of the difference between the transformed correlation coefficients is:

$$\sigma_{(z_1^* - z_2^*)} = \sqrt{\frac{1}{(n_1-3)} + \frac{1}{(n_2-3)}}.$$

Table 11

Reliability Coefficients of Placement Tests by Class

Class	Test	KR-20	n	Mean[a]	S. D.[b]
1989	ALG	0.75	582	22.5	5.8
	ALGA	0.28	582	3.7	1.3
	ALGNA	0.72	582	18.8	5.0
	TRIG	0.70	582	12.4	3.5
	TRIGA	0.21	582	1.1	0.7
	TRIGNA	0.68	582	11.4	3.2
1990	ALG	0.79	692	22.9	6.3
	ALGA	0.33	692	3.7	1.3
	ALGNA	0.76	692	19.2	5.4
	TRIG	0.72	692	12.3	3.6
	TRIGA	0.17	692	1.0	0.7
	TRIGNA	0.71	692	11.2	3.3
1991	ALG	0.79	742	22.9	6.3
	ALGA	0.37	742	3.8	1.3
	ALGNA	0.77	742	19.1	5.5
	TRIG	0.71	742	12.6	3.6
	TRIGA	0.24	742	1.0	0.7
	TRIGNA	0.71	742	11.6	3.3

Note. [a] Mean test score. [b] Standard deviation of the test scores.

Fisher found this transformation to be "almost normal with standard error . . . practically independent of the value of the correlation in the population from which the sample is drawn" (Snedecor & Cochran, 1969, p. 185). Thus, the differences of the values of z^* will be compared to standard normal scores. The differences between the z^* values are provided in Table 12.

Table 12

Tests of Differences Between Test Reliabilities by Class

Classes	Test	$z_1^* - z_2^*$	$\sigma(z_1^* - z_2^*)$	$\frac{z_1^* - z_2^*}{\sigma(z_1^* - z_2^*)}$
1989-1990	ALG	-0.1043	0.05638	-1.8510
	ALGA	-0.0561	0.05638	-0.9948
	ALGNA	-0.1022	0.05638	-1.8132
	TRIG	-0.0489	0.05638	-0.8675
	TRIGA	0.0366	0.05638	0.6498
	TRIGNA	-0.0560	0.05638	-0.9925
1989-1991	ALG	-0.0985	0.05550	-1.7743
	ALGA	-0.0943	0.05550	-1.6999
	ALGNA	-0.1097	0.05550	-1.9759
	TRIG	-0.0199	0.05550	-0.3583
	TRIGA	-0.0380	0.05550	-0.6837
	TRIGNA	-0.0588	0.05550	-1.0588
1990-1991	ALG	0.0059	0.05296	0.1110
	ALGA	-0.0383	0.05296	-0.7224
	ALGNA	-0.0074	0.05296	-0.1404
	TRIG	0.0290	0.05296	0.5481
	TRIGA	-0.0746	0.05296	-1.4083
	TRIGNA	-0.0028	0.05296	-0.0530

In Table 12 none of the KR-20 reliability coefficients for the cognitive subscales were significantly different at the $\alpha = 0.05$ level, after using Bonferroni's correction for three dependent hypothesis tests. Thus, hypotheses H3.2.1, H3.2.2, H3.2.3, and H3.2.4 cannot, in general, be rejected, implying that the reliability for the placement tests, and for the cognitive subscores, had reasonably stable internal consistency measures over the three year span. While the reliability coefficients of the tests were stable, there were strong indications in Table 11 that the levels of reliability of some of the tests were unacceptably low.

While the algebra and trigonometry non-analysis subscales exhibited KR-20 values within the range acceptable for locally developed achievement tests, (e.g., KR-20 > 0.70), the values for the cognitive scales did not. Specifically, the analysis subscales for both content areas, displayed KR-20 values much below what is generally considered acceptable; on the average 0.33 and 0.21 for the respective algebra and trigonometry analysis subtests. This result may be a function of the extremely small number of items composing the subscales.

From the Spearman-Brown prophecy formula (Crocker and Algina, 1986) one can estimate the number of items, like those on the subtests, needed to achieve a KR-20 value of at least 0.70. The formula for performing this calculation is $n = \frac{7(1-r)}{3r}$, where n is the number by which the original number of items must be multiplied in order to achieve the KR-20 value of 0.70, given that the original KR-20 value was r. Table 13 shows the predicted total number of items needed in each subtest to increase its KR-20 value to 0.70.

The data in Table 13 indicate that, on the average, the algebra analysis item subscale needs an additional 24 items to achieve a KR-20 value of at least 0.70 and the trigonometry analysis item subscale needs an additional 17 items. On the average, the non-analysis subscales exceeded the target KR-20 value of 0.70.

The results of this last section cast doubt about the usefulness of the Cognitive model applied to the DFMS placement examinations because of the extremely low reliability of the analysis subscales. The reliability can have a dramatic affect upon the validity coefficients between predictors, analysis subscales scores, and the criterion, final calculus grades. Using classical mental test theory, it is possible to show that the true correlation between a predictor, X, and a criterion, Y, is:

$$\rho_{xy} \leq \sqrt{\rho_{xx'}} \cdot \sqrt{\rho_{yy'}},$$

where ρ_{xy} is the validity coefficient between test X and criterion Y, $\rho_{xx'}$ is the reliability coefficient of test X, and $\rho_{yy'}$ is the reliability coefficient of criterion Y. Thus, it is expected that in this study the analysis subscales scores had a reduced capability to predict final calculus grades because of low reliability and may have contributed to the inconclusive results of the construct validity investigation. The next section will empirically assess the capability of the cognitive variables to predict final calculus grades.

Table 13

Spearman-Brown Calculations for Reliability Coefficients by Test and Class

Class	Test	Original KR-20	Original No. Items	Target KR-20	n	Total No. Items Needed
1989	ALGA	0.28	6	0.70	6.0	36
	TRIGA	0.21	2	0.70	8.8	18
	TRIGNA[a]	0.68	18	0.70	1.1	20
1990	ALGA	0.33	6	0.70	4.7	29
	TRIGA	0.18	2	0.70	10.6	22
1991	ALGA	0.37	6	0.70	4.1	24
	TRIGA	0.24	2	0.70	7.4	15

Note. [a] The TRIGNA subscale for the Classes of 1990 and 1991 exceeded the 0.70 target.

Predictive Validity of Cognitive Model

The predictive validity of the cognitive model variables identified in Question II was investigated, in regard to predicting the final calculus grades, using multiple linear regression. F-tests and t-tests were used to determine if the cognitive variables were statistically significant predictors of final calculus grades. The specific hypotheses tested were:

H3.3.1: Using the regression equation:

$$GRADE = \beta_0 + \beta_1 ALGA + \beta_2 ALGNA + \beta_3 TRIGA + \beta_4 TRIGNA + \beta_5 MATH + \varepsilon \qquad (2),$$

the parameters $\beta_1, \beta_2, \beta_3, \beta_4,$ and β_5 are all simultaneously equal to zero.

H3.3.2: If H3.3.1 is rejected, then some of the parameters β_1, $\beta_2, \beta_3, \beta_4,$ and β_5 in Equation 2 are equal to zero.

In addition, the prediction equations for the three classes were cross-validated. The cross-validations were accomplished by comparing shrinkage values. Shrinkage values of less than 0.10 were interpreted as indicating relatively stable accuracies of predictions across classes.

Table 14 summarizes the results of regressing GRADE onto ALGA, ALGNA, TRIGA, TRIGNA, and MATH for the Classes of 1989, 1990, and 1991. The percentage of variance of GRADE explained by the predictors increased with each succeeding class, from 5.7% in the Class of 1989 to 24.9% in the Class of 1991. In each class, except 1991, the algebra and the trigonometry analysis subscales scores were not significant predictors of GRADE. In all classes, except 1989, the non-analysis subscales scores and the MATH scores were significant predictors of GRADE.

While each of the regression equations explained a statistically significant amount of the variance found in GRADE for each class, the regression equation for the Class of 1989 must be considered as practically ineffective, with a R^2 of 0.06. The other two equations produced reasonable values for R^2, e.g., 0.14 and 0.25.

Table 14

Cognitive Model Regression ANOVA Tables With Parameter Estimates

Class	Model	DF	Residual Sum of Squares	R^2	F	p>F
1989	Cognitive	576	323.4068	0.0571	6.976	0.0001

Parameter Estimates for Cognitive Model Regression

Variable	DF	Parameter Estimate	Standard Error	t for H_0 Parameter = 0	Prob > |t|
Intercept	1	-1.4038	0.3114	4.508	0.0001
ALGA	1	-0.0327	0.0288	-1.138	0.2557
ALGNA	1	0.0057	0.0081	0.707	0.4797
TRIGA	1	0.0587	0.0491	1.193	0.2335
TRIGNA	1	0.0274	0.0116	2.371	0.0181
MATH	1	0.0018	0.00055	3.377	0.0008

Class	Model	DF	Residual Sum of Squares	R^2	F	p>F
1990	Cognitive	686	549.7586	0.1392	22.179	0.0001

Parameter Estimates for Cognitive Model Regression

Variable	DF	Parameter Estimate	Standard Error	t for H_0 Parameter = 0	Prob > |t|
Intercept	1	0.1644	0.3466	0.474	0.6354
ALGA	1	-0.00012	0.0318	-0.004	0.9969
ALGNA	1	0.0413	0.0085	4.861	0.0001
TRIGA	1	0.0175	0.0539	0.324	0.7461
TRIGNA	1	0.0330	0.0125	2.643	0.0084
MATH	1	0.0019	0.00059	3.242	0.0012

Class	Model	DF	Residual Sum of Squares	R^2	F	p>F
1991	Cognitive	736	639.1808	0.2485	48.686	0.0001

Parameter Estimates for Cognitive Model Regression

Variable	DF	Parameter Estimate	Standard Error	t for H0 Parameter = 0	Prob > |t|
Intercept	1	-0.2726	0.3492	-0.781	0.4352
ALGA	1	0.1299	0.0306	4.247	0.0001
ALGNA	1	0.0356	0.0083	4.278	0.0001
TRIGA	1	0.1567	0.0546	2.869	0.0042
TRIGNA	1	0.0553	0.0127	4.350	0.0001
MATH	1	0.0012	0.00062	2.012	0.0446

Thus, the results are evidence that hypothesis H3.3.1 should be rejected for all three classes, i.e., the parameters β_1, β_2, β_3, β_4, and β_5 were not all simultaneously equal to zero. The evidence in Table 14 was mixed for rejecting the hypothesis H3.3.2, that parameters β_1 and β_3 in Equation 2 were equal to zero. There was strong evidence to reject the hypotheses H3.3.3, that β_2, β_4, and β_5 were zero.

The results for H3.3.2 appear to be conditional upon the degree of grade inflation present in the class. When the level of unsuccessful students was less than 11% (see Table 7), not all the cognitive variables were significant predictors of GRADE. But, for the Class of 1991, which had 13.8% unsuccessful cadets, all of the cognitive variables were significant at the $\alpha = 0.05$ level.

Hence, the increasing number of unsuccessful students in the classes is particularly attractive to explain the commensurate pattern of the increasing amount of variance of GRADE accounted for by the cognitive predictors. This apparent increase in the predictability of GRADE by the cognitive predictors was investigated by cross-validation.

Cross-Validation of Cognitive Model

The Cognitive model was cross-validated using the three classes in the study. The regression coefficients obtained from regressing GRADE on ALGA, ALGNA, TRIGA, TRIGNA and MATH for the Class of 1989 were used to predict GRADE using the observed values of the same predictor variables for the Classes of 1990 and 1991. The standard errors of estimate were computed from the difference between the predicted and observed GRADE values. The multiple correlation coefficients between the predicted and observed GRADE values were computed as Pearson-product moment coefficients. This same method was applied using the regression coefficients for the Class of 1990 to predict GRADE for the Class of 1991. The standard error of the estimate and the correlation coefficient were similarly computed. These values are reported in Table 15, while the shrinkage values are displayed in Table 16.

Table 16 shows that the 1991 shrinkage value for the 1989 equation was relatively large, 0.101, indicating that this prediction equation may not produce stable prediction across two years. The 1990 prediction equation seemed to provide an acceptable 1991 shrinkage value, 0.033, and therefore could be expected to produce a stable placement across at least one year. While there is evidence of instability of the accuracies of prediction across the three classes, the effect is not as strong as suggested by the variations in the regression coefficients. This supports the contention that the restriction of range of GRADE across the three classes had a strong influence on the degree of usefulness of the cognitive predictors. Since the analysis subscale scores were significant predictors of GRADE for the Class of 1991, it may be that, in general, the cognitive variables are most useful in predicting final course grades when the number of unsuccessful students are allowed to be around 13% of the class. It is also worthwhile noting that the cognitive variables may become more efficient predictors after increasing the length of the analysis subscales with sufficient numbers of extra items to increase their reliabilities.

Table 15

Cognitive Model Cross-Validation Correlations and Standard Errors[a]

1989 Equation		1990 Cross-validation of 1989 equation		1990 Equation		1991 Cross-validation of 1989 equation		1991 Cross-validation of 1990 equation		1991 Equation	
R	SE	R	SE	R	SE	R	SE	R	SE	R	SE
.239	.749	.320	.920	.373	.895	.397	1.01	.466	.957	.499	.932

Note. [a] Standard error of regression.

Table 16

Cognitive Model Cross-validation Shrinkage Values

Shrinkage(1990,1989) = 0.373 - 0.320 = 0.053
Shrinkage(1991,1989) = 0.499 - 0.397 = 0.102*
Shrinkage(1991,1990) = 0.499 - 0.466 = 0.033

Note. * Unstable prediction, since the value exceeds the 0.10 critical value.

Summary of Cognitive Model Validity

The results for testing Question III can be summarized by noting that there was: (a) acceptable support for the content validity of the cognitive subscales; (b) marginal support for the predictive validity of the cognitive variables with final calculus grades; (c) lack of support of the construct validity of the cognitive subscales; (d) lack of support for acceptable levels of reliability for the analysis subscales.

Two major factors may account for these findings. One factor reported was the likelihood that the small number of analysis items could have caused the low the reliability of the analysis subscales. It was pointed out that the low reliability coefficients necessarily limited the size of the predictive validity coefficients that could be obtained. The other factor was that there was an apparent change in DFMS policies concerning the number of placements in the long and short sequences and the grade distributions of the cadets in these sequences. This researcher posited that the reason for the policy changes may have been in reaction to the excessive inflation of final calculus grades, primarily experienced in the Classes of 1989 and 1990. When the number of unsuccessful cadets reached about 13% of the class, the cognitive variables were both practically and statistically significant, despite the problems of the unreliability of the analysis subscales. Thus, if enough of the right kind of analysis items were added to the analysis subscales, and if the final grades for calculus were controlled to give at least 13% unsuccessful grades, then the cognitive subscales would likely produce acceptable reliability coefficients and all of the cognitive variables would probably be significant predictors of final calculus grades. Despite the marginal validity of the Cognitive model, it is still possible that the model could be useful for placement.

The efficiency of the Cognitive model was compared with that of the Willingham, SOP, and the Computerized placement models. The results of those analyses are reported in the next section.

Question IV Results

Question IV dealt with the efficiency of the Cognitive Model for College Mathematics Placement compared to that of a theoretically based placement model, the Willingham model, and the two empirical placement models used by DFMS, the SOP model and the Computerized placement model. This section reports the results of the comparisons of the various models. Two methods were used for the comparisons, hit-and-miss tables and generalized *F*-tests.

Comparing the Cognitive and Willingham Models

The Cognitive and Willingham models were applied to each of the three classes of cadets; the resulting hit-and-miss tables were compared, and the generalized *F*-tests were evaluated. Before the comparisons were performed; however, the Willingham model was first validated with methods similar to those used to validate the Cognitive model.

Validating the Willingham Model

The validity of the Willingham model, as applied to the Classes of 1989, 1990, and 1991, was investigated through an analysis of the content and predictive validity of the algebra and trigonometry placement tests. The methods used to analyze the content and predictive validity of the model were the same as those previously used for the Cognitive model. Thus, multiple linear regression was used to assess the capability of ALG and TRIG, as specified by the Willingham model, to predict final calculus grades. The content validity of the two placement tests was assessed using the percentage of agreement and the average Likert scale value on the placement test items obtained through a content validity questionnaire (Appendix B).

Content validity of overall placement tests.

The content validity of the placement tests was investigated by having the precalculus course director and two precalculus instructors respond to a single Likert scaled item on the content validity questionnaire. The Likert item required each judge to classify each of the placement test items as one that does not test a content topic in the syllabus, or as either a poor, adequate, good, or an excellent test item of a content topic in the syllabus (see Appendix B). These responses were coded as 1 through 5 respectively. The test was considered valid if the average coded value was at least 2.5 and the average percentage of agreement among the judges was at least 67%. The results of the content validity assessment are listed in Table 17.

Table 17
Willingham Model Content Validity Agreement

Test	No. of Items	Average Percent	Modal Judgment	Average Judgment
ALG	40	65.0	4	4.0
TRIG	20	75.0	4	4.1

As reported in Table 17, only the algebra placement test did not meet the content validity criterion. Even though the average coded score of 4.0 exceeded the criterion of 2.5, the average percentage of agreement, 65.0%, was smaller than the required 67%. Appendix B shows that there were five items which had 0% agreement, but all of these items had score patterns of coded judgments of 3, 4, 5. Thus, while a true disagreement existed among the judges, the disagreement was not about the items not being representative of the content of the course nor that the items were poor test items. All of the judges felt that these five items were at least adequate test items. Without these five average coded scores the average percent agreement was 74.5 %. Thus, the algebra test was composed of items which were all judged to be at least adequate test items of the content of the precalculus course.

Reliability of placement tests.

This section reports the results of investigating the internal consistency of the placement tests for adequacy and stability over time. The statistical hypotheses tested in this section were:

> H4.1.1: The reliability coefficients of the algebra placement tests are not pairwise different for the classes of 1989, 1990, and 1991.
>
> H4.1.2: The reliability coefficients of the trigonometry placement tests are not pairwise different for the classes of 1989, 1990, and 1991.

The KR-20 reliability coefficient was used to determine the internal consistency of each of the placement tests for each class of cadets. These values were then compared using Fisher's z^* statistics to test for significant differences between the classes. Table 18 displays the KR-20 values for each placement test by class. The table indicates that each of the tests produced a KR-20 which exceeded 0.70, an acceptable reliability coefficient for a locally constructed test.

The stability of the KR-20 values across the three classes were analyzed by transforming the KR-20 values into Fisher z^* values and testing the significant differences between the classes. The differences between the z^* values are reported in Table 19.

Table 19 indicates that the KR-20 coefficients for the algebra and trigonometry placement tests were <u>not</u> significantly different at the $\alpha = 0.05$ level, after using Bonferroni's correction for three dependent hypothesis tests. Thus, hypothesis H4.1.1 may not be rejected, implying that the algebra placement test had stable internal consistency measures over the three-year span. There is also <u>no</u> evidence to reject hypothesis H4.1.2; hence, the reliability for the trigonometry placement test _was_ assumed to be consistent over the three classes of cadets.

Thus, it is expected that the reliability coefficients did not overly limit the predictive validity coefficients of the placement tests with final calculus grades. The next section reports the analysis of the predictive validity of the Willingham model.

Table 18

Reliability Coefficients of Placement Tests by Class

Class	Test	KR-20	n	Mean[a]	S.D.[b]
1989	ALG	0.75	582	22.5	5.8
	TRIG	0.70	582	12.4	3.5
1990	ALG	0.79	692	22.9	6.3
	TRIG	0.72	692	12.3	3.6
1991	ALG	0.79	742	22.9	6.3
	TRIG	0.71	742	12.6	3.6

Note. [a] Mean test score. [b] Standard deviation of scores on the placement test.

Table 19

Tests of Differences Between Test Reliabilities by Class

Classes	Test	$z_1^* - z_2^*$	$\sigma(z_1^* - z_2^*)$	$\frac{z_1^* - z_2^*}{\sigma(z_1^* - z_2^*)}$
1989-1990	ALG	-0.1043	0.05638	-1.8510
	TRIG	-0.0489	0.05638	-0.8675
1989-1991	ALG	-0.0985	0.05550	-1.7743
	TRIG	-0.0199	0.05550	-0.3583
1990-1991	ALG	0.0059	0.05296	0.1110
	TRIG	0.0290	0.05296	0.5481

Predictive validity of Willingham model.

The predictive validity of the Willingham model was investigated for each class, in regard to predicting the final calculus grades, using multiple linear regression. F-tests and t-tests were used to determine if the variables of the Willingham model were significant predictors of final calculus grades. The specific hypotheses tested were based on the regression equation:

$$GRADE = \beta_0 + \beta_1 ALG + \beta_2 TRIG + \varepsilon \quad (3).$$

The specific hypotheses tested were:

H4.2.1: The parameters β_1 and β_2 of Equation 3 are both simultaneously equal to zero.

H4.2.2: If H4.1.1 is rejected, then at least one of the parameters β_1 and β_2 in Equation 3 is equal to zero.

Table 20 contains the analysis of variance tables and parameter estimates for the Willingham model applied to the data from each of the classes. Table 20 provides strong evidence to reject H4.2.1, that the parameters β_1 and β_2 are simultaneously equal to zero for each of the classes. Additionally, the parameter estimates listed in Table 20 indicate that the algebra and trigonometry total placement scores are generally significant predictors of the final grades in calculus. The only exception to this was that ALG was not a significant predictor of final grades in calculus for the Class of 1989. However, ALG became a significant predictor as the number of unsuccessful cadets increased for the subsequent classes. Thus, in general, there was evidence to reject H4.2.2, that the parameters β_1 and β_2 were equal to zero for each of the classes.

Table 20 implies that the Willingham model is valid for predicting final course grades but, these data do not speak to the stability of the prediction. The stability issue was analyzed by the cross-validation of the individual Willingham model prediction equations.

Table 20

Willingham Model Regression ANOVA Tables With Parameter Estimates

Class	Model	DF	Residual Sum of Squares	R^2	F	p>F
1989	Willingham	579	330.8437	0.0354	10.628	0.0001

Parameter Estimates for Willingham Model Regression

Variable	DF	Parameter Estimate	Standard Error	t for H_0 Parameter = 0	Prob > |t|
Intercept	1	2.3081	0.1394	16.553	0.0001
ALG	1	0.0088	0.0062	1.417	0.1570
TRIG	1	0.0324	0.0102	3.163	0.0016

Class	Model	DF	Residual Sum of Squares	R^2	F	p>F
1990	Willingham	689	559.5400	0.1238	48.695	0.0001

Parameter Estimates for Willingham Model Regression

Variable	DF	Parameter Estimate	Standard Error	t for H_0 Parameter = 0	Prob > |t|
Intercept	1	1.1797	0.1429	8.253	0.0001
ALG	1	0.0440	0.0064	6.904	0.0001
TRIG	1	0.0274	0.0110	2.485	0.0132

Class	Model	DF	Residual Sum of Squares	R^2	F	p>F
1991	Willingham	739	653.2467	0.2320	111.624	0.0001

Parameter Estimates for Willingham Model Regression

Variable	DF	Parameter Estimate	Standard Error	t for H_0 Parameter = 0	Prob > |t|
Intercept	1	0.3608	0.1453	2.483	0.0132
ALG	1	0.0547	0.0064	8.495	0.0001
TRIG	1	0.0675	0.0113	5.983	0.0001

Cross-Validation of Willingham model.

The Willingham model was cross-validated with the same methods used to cross-validate the Cognitive model. Regression coefficients, obtained from regressing GRADE on ALG and TRIG, for one class were used to predict GRADE for the subsequent classes. Multiple correlation coefficients and standard errors of estimate were computed from these regressions. These values are provided in Table 21. Shrinkage values were also computed and reported in Table 22. Shrinkage values in excess of 0.10 were considered as evidence of unstable predictions.

Table 21
Willingham Model Cross-Validation Correlations and Standard Errors

1989 Equation		1990 Cross-validation of 1989 equation		1990 Equation		1991 Cross-validation of 1989 equation		1991 Cross-validation of 1990 equation		1991 Equation	
R	SE	R	SE	R	SE	R	SE	R	SE	R	SE
.188	.756	.315	.925	.352	.901	.461	1.01	.475	.958	.482	.940

The multiple correlation coefficients for the Classes of 1990 and 1991 are typical for prediction models as discussed in the review of literature. The multiple correlation coefficient for the Class of 1989 is extremely low, which may be explained by the restriction of range of GRADE. Table 21 shows the steady increase of the correlations of ALG and TRIG with GRADE, as discussed previously. Interestingly, the multiple correlation coefficients for the 1989 prediction equation applied to the 1990 and 1991 data are similar to those obtained from the least squares estimates. The degree of similarity is emphasized with the shrinkage values which are listed in Table 22.

Table 22

Willingham Model Cross-validation Shrinkage Values

Shrinkage(1990,1989) = .352 - .315 = .037
Shrinkage(1991,1989) = .482 - .461 = .021
Shrinkage(1991,1990) = .482 - .475 = .007

Table 22 shows that all the shrinkage values are below the critical value of 0.10. Thus, each of the prediction equations provide an acceptable level of consistency in prediction. The apparent change of the accuracy in the prediction over the years, i.e., the increase of $\underline{R}$, may not be due primarily to improved estimates of the prediction parameters. While it is true that the 1991 prediction equation explains more of the variance in the 1991 data than the 1989 equation explained in the 1989 data, the 1989 prediction equation explains nearly the same amount of the variance of the 1991 data as did the 1991 equation. The least squares estimates in the 1991 prediction equation explain the maximum amount of the variance in the final course grades that predictor variables can explain for the 1991 data. This supports the suspicion that the low level of prediction of the 1989 prediction equation of the 1989 data may be more a function of the restriction of range in GRADE rather than capability of the variables ALG and TRIG to predict GRADE.

Summary of Willingham Model Validity

Overall, the Willingham model of placement seems to be valid in that the instruments producing the scores are acceptably reliable, produce reasonable correlations with final course grades in calculus, and show stable levels of accuracy of prediction. The predictive validity for the Willingham model was low for the Classes of 1989 and 1990, but the evidence pointed to a changing environment where the policy of assigning grades became more restrictive, which affected the number of unsuccessful grades assigned. It was seen that as the grade inflation was reduced, the restriction of range on GRADE decreased, increasing the accuracy of prediction. Therefore, efficiencies of placements from the Willingham model may reasonably be compared to the Cognitive and other empirical models. The first method of comparison used hit-and-miss tables.

Cognitive and Willingham Models Hit-and-Miss Table Comparisons.

The hit-and-miss tables for each class, based on both models, were constructed from hypothetical placements accomplished in accordance with the methods of Appenzellar and Kelley (1983). These methods required that Kelley Tables (Appendix D) be constructed to develop a set of candidate cutoff values to be used for the hypothetical placements. The following guidelines were used to suggest possible cutoff values of the Composite Score (the predicted final course grade from the Cognitive or Willingham model regression equation):

1. The Expected Composite Score for those students with a Final Course Grade of 2.00.

2. The Composite Score for those students with an Expected Final Course Grade of 2.00.

3. The Composite Score for which the percentages of errors of students in each academic performance category (Satisfactory or Unsatisfactory), were most nearly equal.

4. The Composite Score for which the overall percentages of errors were most nearly equal.

5. The Composite Score that would have cut off, or held back, approximately the same number of students as were in the Unsatisfactory performance category.

6. The Composite Score that would have maximized the overall accuracy of placement.

Table 23 contains the candidate cutoff Composite Scores from applying the six guidelines to the various classes of cadets.

Table 23

Candidate Composite Score Cutoffs by Guideline, Class, and Placement Model

	Guideline					
Class/Model	1	2	3	4	5	6
1989/Willingham	2.88	2.00	2.90	2.60	2.57	2.50
1990/Willingham	2.46	2.00	2.40	2.10	2.10	1.40
1991/Willingham	2.36	2.00	2.30	1.90	1.87	1.50
1989/Cognitive	2.86	2.00	2.80	2.50	2.53	2.40
1990/Cognitive	2.45	2.00	2.00	2.10	2.06	1.40
1991/Cognitive	2.35	2.00	2.30	1.90	1.89	1.30
1991[a]/Cognitive	2.31	2.00	2.20	1.90	1.86	1.30

Note. [a] The Class of 1991 without hand-placed cadets.

A Composite Score of 2.00 was selected as the cutoff score to perform the hypothetical placements for all placement models for each class of cadets. This score was selected because it is the lowest predicted final course grade considered successful. Also, a Composite Score of 2.00 is close to other Composite Scores which produce placements which will give about the same number of overall errors, i.e., too high and too low, and that cut off the same number of cadets which were observed to be in the unsuccessful category. Thus, 2.00 may be considered a satisfying cutoff from both theoretical aspects and empirical evidence. The other candidate cutoff Composite Scores were considered practically infeasible.

The cutoff 2.00 was used to perform hypothetical placements for each class. These placements were in turn analyzed in terms of the number of hits (correct placements) and misses (incorrect placements) according to the framework established in Table 4 in the Methodology Chapter. Table 4 is reproduced here for ease of interpreting the results. Table 24 displays the individual hit-and-miss tables for each class by placement model. The total number of hits and misses are summarized in Table 25.

Table 4

General Hit-and-Miss Table Interpretation

Hypothetical Placement	Actual Placement			
	Long Sequence		Short Sequence	
	Performance		Performance	
	U[a]	S[b]	U	S
Long Sequence	Correct[c]	Unknown	Correct	Incorrect
Short Sequence	Incorrect	Unknown	Incorrect	Correct

Note. [a] Unsatisfactory observed grade in calculus. [b] Satisfactory observed grade in calculus. [c] Correctness of hypothetical placement.

Table 25 indicates that the Willingham model consistently produced more correct placements than the Cognitive model. However, the actual differences between the numbers (and percents) of hits, 0 (0%), 5 (0.7%), and 8 (1.1%) for the Classes of 1989, 1990,

Table 24
Individual Hit-and-Miss Tables by Placement Model and Class

Cognitive Model Class of 1989	Actual Placement			
	Long Seq.		Short Seq.	
	Performance		Performance	
Hypothetical Placement	U	S	U	S
Long Sequence	0	0	0	0
Short Sequence	4	117	7	454

Cognitive Model Class of 1990	Actual Placement			
	Long Seq.		Short Seq.	
	Performance		Performance	
Hypothetical Placement	U	S	U	S
Long Sequence	8	32	3	12
Short Sequence	14	69	48	506

Cognitive Model Class of 1991	Actual Placement			
	Long Seq.		Short Seq.	
	Performance		Performance	
Hypothetical Placement	U	S	U	S
Long Sequence	42	71	8	22
Short Sequence	18	54	35	492

Cognitive Model Class of 1991 Reduced	Actual Placement			
	Long Seq.		Short Seq.	
	Performance		Performance	
Hypothetical Placement	U	S	U	S
Long Sequence	40	67	10	19
Short Sequence	15	39	30	404

Willingham Model Class of 1989	Actual Placement			
	Long Seq.		Short Seq.	
	Performance		Performance	
Hypothetical Placement	U	S	U	S
Long Sequence	0	0	0	0
Short Sequence	4	117	7	454

Willingham Model Class of 1990	Actual Placement			
	Long Seq.		Short Seq.	
	Performance		Performance	
Hypothetical Placement	U	S	U	S
Long Sequence	7	31	1	4
Short Sequence	15	70	50	514
Willingham Model Class of 1991	Actual Placement			
	Long Seq.		Short Seq.	
	Performance		Performance	
Hypothetical Placement	U	S	U	S
Long Sequence	45	72	7	16
Short Sequence	15	53	36	498

Table 25
Summary of Hit-and-Miss Tables

Class	Model	Hit n	Hit (%)	Miss n	Miss (%)	Unknown n	Unknown (%)
1989	Cognitive	454	(78.0)	11	(1.9)	117	(20.1)
	Willingham	454	(78.0)	11	(1.9)	117	(20.1)
	SOP	458	(78.7)	7	(1.2)	117	(20.1)
1990	Cognitive	517	(74.7)	74	(10.7)	101	(14.6)
	Willingham	522	(75.4)	69	(10.0)	101	(14.6)
	SOP	540	(78.0)	51	(7.9)	101	(14.6)
1991	Cognitive	542	(73.0)	75	(10.1)	125	(16.8)
	Willingham	550	(74.1)	67	(9.0)	125	(16.8)
	Computer[a]	574	(77.4)	43	(5.8)	125	(16.8)
1991	Cognitive Reduced	454	(72.8)	64	(10.3)	106	(17.0)
	Computer[b] Reduced	478	(76.6)	40	(6.4)	106	(17.0)

Note. [a] Computerized model hit-and-miss performance. [b] Computerized model hit-and-miss performance with hand-placed cadets removed.

and 1991 respectively, were considered negligible by this researcher because of the small magnitudes. Generalized F-tests were performed to gain a consensus of the results of the hit-and-miss table comparisons.

Cognitive and Willingham Models Generalized F-Test Comparisons

Generalized *F*-tests were performed using data from the ANOVA tables from the Cognitive and Willingham models' regressions to test the following hypothesis:

H4.3: In Equation 2, $\beta_1 = \beta_2$, $\beta_3 = \beta_4$, and $\beta_5 = 0$.

A significant *F* value was interpreted as evidence indicating that H4.3 should be rejected and that the Cognitive model is a more efficient prediction model than the Willingham model. Otherwise, H4.3 could not be rejected, meaning that the Willingham model is as efficient a prediction model as the Cognitive model. The results of the generalized *F*-tests are presented in Table 26.

Each of the generalized *F*-tests was significant thus hypothesis H4.3 should be rejected; that is, that $\beta_1 = \beta_2$, $\beta_3 = \beta_4$, and $\beta_5 = 0$ for all three classes. While the generalized *F*-test implies that the Cognitive model consistently explained more of the variance in the final calculus grades for each of the classes, one may notice that the practical improvement is small. The practical improvement is demonstrated by the differences 2.17%, 1.54%, and 1.65% for the Classes of 1989, 1990, and 1991 respectively. It is likely that the differences were statistically significant because of the large number of degrees of freedom in the regressions.

Summary of Cognitive and Willingham Models Comparison

The results of the comparisons between the Cognitive and Willingham models were mixed. The hit-and-miss tables showed that the Willingham model consistently produced more correct placements, with an average increase of 4.3 (0.6%). The generalized *F*-tests, however, showed that the Cognitive model consistently produced more accurate predictions of final course grade, with an average increase of 1.8% more of the variance in the final calculus grades explained. However, there is little practical difference between the efficiencies of the two models using either measure since the magnitudes of the differences in the number of correct placements and amount of variance explained with the prediction equations are so small.

Table 26

Cognitive vs. Willingham Models Generalized F-Tests Analysis of Variance Tables

Class	Model	DF	Residual Sum of Squares	R^2	*F*	*p*>*F*	General *F*
1989	Cognitive	576	323.4068	0.0571	6.976	0.0001	4.415**
	Willingham	579	330.8437	0.0354	10.628	0.0001	
1990	Cognitive	686	549.7586	0.1392	22.179	0.0001	4.068**
	Willingham	689	559.5400	0.1238	48.695	0.0001	
1991	Cognitive	736	639.1808	0.2485	48.686	0.0001	5.399**
	Willingham	739	653.2467	0.2320	111.624	0.0001	

** $p < 0.01$

This last section showed that the two learning theory based placement models were, in general, equally effective. In the next two sections, the Cognitive model will be compared with the two empirical placement models used by DFMS to perform the actual placements for the Classes in this study.

Cognitive and SOP Models Comparison

The Cognitive model was compared with the SOP model using the Classes of 1989 and 1990; the Class of 1991 was not used since those cadets were placed using the Computerized Placement model. The comparisons were accomplished with hit-and-miss tables and generalized *F*-tests. The hit-and-miss tables were constructed from the hypothetical placements using the Cognitive model, as well as the observed placements actually performed by the SOP model. The total number of hits and misses was extracted from these tables and compared. The number of hits for the Cognitive model was found using the guidelines established in Table 4; that is, the sum of the number of cadets in the categories denoted as correct. The number of hits for the SOP model was found by summing the number of cadets who were placed into the long sequence and were unsuccessful in calculus and the number of cadets who were placed into the short sequence and were successful in calculus. The individual hit-and-miss tables were presented in Table 24 and summarized in Table 25. The portion of Table 25 relevant to the Cognitive and SOP models comparison was extracted and displayed in Table 27.

Table 27 reports modest increases in efficiencies by the SOP model over the Cognitive model, with increases of 4 (0.7%) and 23 (3.3%) hits for the Classes of 1989 and 1990 respectively. While 0.7% increase in efficiency is not practically significant, the 3.3% improvement may be considered practically significant. It may be there is little difference, in terms of number of hits-and-misses of placements between the placement models when there will be few unsuccessful students. Under conditions where there will likely be nearly 10% unsuccessful cadets, the SOP model may be more efficient. However, The effects of the large number of "Unknowns" in the Cognitive model hit-and-miss tables are greater than the observed differences, so the results are inconclusive.

Table 27
Cognitive vs. SOP Models Hit-and-Miss Tables

		Hit		Miss		Unknown	
Class	Model	n	(%)	n	(%)	n	(%)
	Cognitive	454	(78.0)	11	(1.9)	117	(20.1)
1989	SOP	458	(78.7)	7	(1.2)	117	(20.1)
	Cognitive	517	(74.7)	74	(10.7)	101	(14.6)
1990	SOP	540	(78.0)	51	(7.9)	101	(14.6)

This section showed that one of the empirical placement models tended to be more efficient than the Cognitive model. It remains to be seen if the other empirical placement model does as well. The next section reports the comparison of the empirically based Computerized Placement model with the Cognitive model.

Cognitive and Computerized Models Comparison

The comparison between the Cognitive and Computerized models used a generalized F-test and hit-and-miss tables. These procedures were applied only to the Class of 1991, since that class was placed using the Computerized model. The whole Class of 1991 was not used for the comparison since there were a large number of cadets who were hand-placed, i.e., not placed in accordance with the Computerized model. Thus, 118 hand-placed cadets were removed from the analysis.

Cognitive and Computerized Models Hit-and-Miss Table Comparison.

As before, the hit-and-miss table was constructed from a hypothetical placement based on the methods of Appenzellar and Kelley (1983). The Kelley Tables (Appendix D) were developed in order to obtain a set of candidate cutoff Composite Scores (see Table 23). The Composite Scores were the Expected Final Calculus Grades derived from applying Equation 3 to the reduced Class of 1991. Consistent with the previous analyses, the Composite Score of 2.00 was selected to perform the hypothetical placement. An individual hit-and-miss table was constructed (see Table 24), for the Cognitive model, which was summarized in Table 28. The Computerized model was used to perform the actual placement, and the number of hits and misses was computed in the same way as was done for the SOP hit-and-miss tables. The total number of hits and misses was also summarized in Table 28.

Table 28
Cognitive vs. Computerized Models Hit-and-Miss Table

		Hit		Miss		Unknown	
Class	Model	n	(%)	n	(%)	n	(%)
1991	Cognitive Reduced	454	(72.8)	64	(10.3)	106	(17.0)
	Computerized Reduced	478	(76.6)	40	(6.4)	106	(17.0)

Table 28 shows that the Computerized model produced more correct placements than the Cognitive model with an increase of 24 (3.8%) correct placements. The magnitude appears to be practically significant. A generalized $\underline{F}$-test was performed to gain a consensus of the results of the hit-and-miss table comparison.

Cognitive and Computerized Models Generalized F-Tests Comparison

The purpose of this section is to report the results of investigating hypotheses:

H4.2.1: In Equation 6, $\beta_2 = \beta_3 = \beta_6 = \beta_7 = \beta_8 = 0$.

H4.2.2: If H4.2.1 is rejected, then some of the $\beta_2, \beta_3, \beta_6, \beta_7$, and β_8 in Equation 6 are equal to zero.

H4.2.3: In Equation 6, $\beta_2 = \beta_3, \beta_6 = \beta_7$, and $\beta_8 = 0$,

where Equation 6 is defined in the paragraph below.

These hypotheses were designed to compare the efficiency of the Cognitive and Computerized models with a generalized F-test and associated t-tests from two multiple linear regressions. These regressions were computed for each model where the actual sequence of the cadets from the reduced Class of 1991 were regressed onto the appropriate set of predictors. The Computerized model regression was of the form:

$$PASS = \beta_0 + \beta_1 CALC2 + \beta_2 ALG + \beta_3 ACI + \beta_4 CALC1 + \beta_5 TRIG + \beta_6 CALC3 + \varepsilon \quad (5).$$

The Cognitive model used in this comparison represented an expanded version of Equation 5, where the cognitive variables ALGA, ALGNA, TRIGA, and TRIGNA replaced ALG and TRIG, with the cognitive variable MATH added. Thus, the Augmented Cognitive regression model took the form:

$$PASS = \beta_0 + \beta_1 CALC2 + \beta_2 ALGA + \beta_3 ALGNA + \beta_4 ACI + \beta_5 CALC1 + \beta_6 TRIGA + \beta_7 TRIGNA + \beta_8 MATH + \beta_9 CALC3 + \varepsilon \quad (6).$$

The results of the generalized F-test is presented in Table 29, along with the parameter estimates for the Augmented Cognitive model.

The analysis of variance table shows that both of the regression equations provide a significant prediction of the successfulness of cadets in calculus. Hence, the evidence suggests that hypothesis H4.2.1 should be rejected; that in Equation 6, $\beta_2 \neq \beta_3 \neq \beta_6 \neq \beta_7 \neq \beta_8 \neq 0$. The parameter estimates of the Augmented Cognitive model indicate that the analysis variables, ALGA and TRIGA, are the only cognitive variables which were significant predictors of cadet success in calculus. These data suggest that ALGA and TRIGA most likely will be useful in defining a discriminant function. Thus, the evidence appears to indicate that, in Equation 6, $\beta_2 \neq 0$ and $\beta_6 \neq 0$. Additionally, the evidence implies that $\beta_3 = 0$, $\beta_7 = 0$, and $\beta_8 = 0$.

Finally, the observed $\underline{F}$ value from the generalized $\underline{F}$-test, 3.122, must be compared with $\underline{F}(\alpha = 0.05, 3, 614)$ which is about 2.62. Thus, the hypothesis, H4.2.3, should be rejected; i.e., in Equation 6, $\beta_2 \neq \beta_3, \beta_6 \neq \beta_7$, and $\beta_8 \neq 0$. That is, the Augmented Cognitive model discriminated between the successful and unsuccessful groups of cadets significantly better than the less complicated Computerized model.

The results of this section were mixed. The evidence showed that while the hit-and-miss table results favored the efficiency of the Computerized model, the generalized $\underline{F}$-test and $\underline{t}$-tests supported the efficiency of the Cognitive model. Regression Equation 6 indicated that the Computerized model should be augmented with analysis cognitive variables in order to improve the discrimination between successful and unsuccessful cadets in calculus. The other cognitive variables did not seem to improve the discrimination.

Table 29

Cognitive vs. Computerized Models Generalized F-Tests Analysis of Variance Table

Model	DF	Residual Sum of Squares	R^2	F	p>F	General F
Augmented	614	64.0368	0.2049	17.578	0.0001	3.122*
Computerized	617	65.0137	0.1927	24.553	0.0001	

Parameter Estimates for Augmented Cognitive Model Regression

Variable	DF	Parameter Estimate	Standard Error	t for H_0: Parameter = 0	Prob > |t|
Intercept	1	-0.0468	0.1838	-2.545	0.0112
CALC2	1	0.0038	0.0043	0.875	0.3819
ALGA	1	0.0313	0.0115	2.726	0.0066
ALGNA	1	0.0045	0.0032	1.426	0.1545
ACI	1	0.00031	0.000068	4.653	0.0001
CALC1	1	0.0097	0.0021	4.659	0.0001
TRIGA	1	0.0547	0.0209	2.612	0.0092
TRIGNA	1	0.0033	0.0051	0.641	0.5220
CALC3	1	0.0103	0.0206	0.499	0.6182
MATH	1	-0.000043	0.00025	-0.169	0.8659

* $p < 0.05$

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

Summary of Findings

This study developed a college mathematics placement model based on cognitive science learning theory, identified prediction variables consistent with that model, validated the model using historical placement data, and compared the effectiveness of the model with other theoretical and empirical models. The Cognitive Model of College Mathematics Placement proved to be a synthesis of Skemp's (1979) notions about the schemas of successful mathematics learners and Wilson's (1971) technology for assessing various classifications of mathematical behavior.

Specifically, the Cognitive model is based on the assumption that students' who possess schemas (a) with a high degree of agreement, in terms of accuracy and completeness with the mathematics content requirements of a prerequisite course; and (b) which contain a large number of connections between the different prerequisite schemas, were likely to succeed in learning the content of a subsequent mathematics course. It was argued that novel, or analysis level, mathematics test items assess the degree of connectedness of students' schemas and that non-analysis items assess the degree of accuracy and completeness of the prerequisite schemas. The Cognitive model maintains that a placement test composed of appropriate analysis and non-analysis items, developed from a conceptual analysis of the prerequisite and criterion courses, is useful to predict students' success or failure in the criterion course. Other measures also were determined to enhance the assessment of the completeness of students' precalculus schemas; the SAT-M and the ACT-M are two examples. The model calls for the placement of students to be accomplished using a cutoff score identified with the methods of Appenzellar and Kelley (1983). The validation of a cognitive placement system is best achieved through trait-treatment interaction methods. Alternative methods require investigating various aspects of validity.

The Cognitive model was validated by applying the model to the algebra and trigonometry placement tests response data from the U.S. Air Force Academy graduating classes of 1989, 1990, and 1991. Items on the algebra and trigonometry placement tests were categorized using the four major cognitive levels of Wilson's (1971) Model of Mathematics Achievement. These responses formed sets of either analysis or non-analysis items, as determined by combining the responses from seven expert judges on a construct validity questionnaire. The judges classified six out of the 40 algebra placement test items and two out of the 20 trigonometry placement test items as analysis items; these sets of items formed two "analysis" subtests. The remainder of the items were classified as non-analysis items and constituted the two "non-analysis" subtests. Confirmatory factor analyses were performed on the individual cadet responses to empirically confirm the hypothesized analysis/non-analysis structure using a Procrustes oblique rotation. It was hypothesized that the factor pattern of the cadets' responses would correlate with the analysis/non-analysis classification made by the expert judges. There was no evidence to support this hypothesis for any of the classes.

The second form of validity investigated was content validity. The content validity of the cognitive subtests was investigated using a questionnaire containing each of the placement test items. Three expert judges responded to a five-choice Likert item for each placement test item. The judges' responses assessed each item in terms of its degree of appropriateness as a test of the content in the precalculus course. There was strong evidence that the trigonometry subtests and the algebra non-analysis subtest were content valid; however, the judges disagreed about content validity of the algebra analysis subtest. The disagreements did not concern the relevance of the items to the precalculus syllabus, rather the disagreements focused on whether the items were adequate, good, or excellent test items. All of the items were considered to be at least adequate test items for the content presented in the precalculus course.

The final form of validity investigated was predictive validity. The predictive validity of the cognitive subtests was investigated using multiple linear regression, where the dependent variable was final calculus grades (GRADE) and the independent variables were the algebra and trigonometry analysis subtest scores (ALGA and TRIGA), the algebra and trigonometry non-analysis subtest scores (ALGNA and TRIGNA), and the equipercentile-equated SAT-M and ACT-M scores (MATH). It was hypothesized that each of these cognitive variables would be a significant predictor of GRADE. The data indicated that different cognitive variables were significant predictors for different classes. TRIGNA and MATH were the only cognitive predictor variables significant at the $\alpha = 0.05$ level in the Class of 1989; however, ALGNA, TRIGNA, and MATH were all significant in the Class of 1990, and all of the cognitive variables were significant in the Class of 1991.

A major factor contributing to this phenomenon was the restriction of range on GRADE. It was suspected that the Classes of 1989 and 1990 experienced grade inflation, because only 1.9% in the Class of 1989 received less than a C grade in calculus while 10.5% and 13.9% were unsuccessful in the Classes of 1990 and 1991, respectively. This phenomenon apparently affected the corresponding quality of the predictions, with the Cognitive model explaining 5.7%, 13.9%, and 24.9% of the variance in the calculus grades for the

Classes of 1989, 1990, and 1991, respectively. The data also showed that the accuracy of prediction was not stable over the three classes, because the shrinkage value between the 1989 and 1991 prediction equations was relatively large. The 1990 prediction equation applied to the Class of 1991 data, however, provided an $\underline{R}$ value similar to the least squares estimate for the same data.

The reliability of the analysis subtests was unacceptably low, which almost certainly had an adverse impact on the assessment of the construct and predictive validity of the Cognitive model. The KR-20 reliability coefficient was determined for each subtest for each class and compared across the three classes. It was hypothesized that there would be no significant differences between the coefficients for comparable tests across the classes. The evidence did not support the rejection of this hypothesis. The reliability coefficients, on the average 0.33 and 0.21 for the respective algebra and trigonometry analysis subtests, were very low. This result was probably because of the small number of items in each of the subtests. Thus, the Spearman-Brown prophecy formula was used to estimate the number of additional items that would be needed to achieve a KR-20 coefficient of 0.70. Approximately 24 algebra analysis items and 17 trigonometry analysis items should be added to achieve the target reliability coefficient. The low reliability coefficients of the analysis subtests very likely limited the capability of ALGA and TRIGA to predict GRADE. It is reasonable to expect that the same effect also reduced the capability of the confirmatory factor analysis to verify the construct validity.

Overall, the results of the validation study of the Cognitive model applied to the placement tests at the Academy demonstrated that there were: (a) acceptable support for the content validity of the cognitive subscales; (b) marginal support for the predictive validity of the cognitive variables with final calculus grades; (c) lack of empirical support for the construct validity of the cognitive subscales; (d) unacceptable levels of reliability for the analysis subscales.

Further research was conducted to determine the relative effectiveness of the Cognitive model compared to another learning theory-based placement model, the Willingham model, and two empirical placement models, the SOP model and the Computerized model. The comparisons were performed using hit-and-miss tables and generalized F-tests.

Before the comparisons were made, the Willingham model was validated with techniques similar to those used to validate the Cognitive model. The Willingham model of placement was determined to be valid because the total algebra and trigonometry placement tests were acceptably content valid and had reasonable levels of reliability, produced moderate correlations with final course grades in calculus, and showed stable accuracy of prediction. The predictive validity for the Willingham model was low for the Classes of 1989 and 1990, as was also true for the Cognitive model.

Hit-and-miss tables were constructed after hypothetical placements were performed using cutoff scores developed from the methods of Appenzellar and Kelley (1983). A predicted final calculus grade of 2.00 from each of the theory-based models was used as a cutoff for the hypothetical placements. Those cadets with predicted final calculus grades of at least 2.00 were hypothetically placed into the short sequence, calculus only, while cadets with predicted final calculus grades of less than 2.00 were hypothetically placed into the long sequence, precalculus-calculus.

The number of hypothetical placement hits-and-misses were determined according to the scheme described in Table 4. The number of hits-and-misses for the empirical models was taken directly from the data, as the Classes of 1989 and 1990 were actually placed using the SOP model and the Class of 1991 was actually placed using the Computerized model. The hit-and-miss tables demonstrated that both the Willingham model and the empirical models consistently produced more correct placements than did the Cognitive model; on the average, 4.3 (0.6%) and 17.0 (2.6%) respectively. However, the increase in the number of correct placements was not considered practically significant because it was so small.

The generalized *F*-tests consistently indicated that the Cognitive model fit the data better than did either the Willingham or the Computerized models. While the Cognitive model was shown to explain more of the variance in the data, significant at least at the $\alpha = 0.05$ level, the actual increase of variance explained was on the average 1.6%; thus, the practical improvement provided by the Cognitive model was questionable. This apparent contradiction was explained by noting that the results of the analyses of the hit-and-miss tables possessed a substantial margin of error because of the large number of placements classified into the "Unknown" category (see Table 25). The results of the generalized *F*-tests may be closer to reality since this technique did not depend on the classification of the placements.

Conclusions

The findings of this research demonstrated that:

1. A Cognitive Model of College Mathematics Placement could be developed from Skemp's cognitive science learning theory and Wilson's Model of Mathematical Achievement.

2. Locally developed placement examinations are the best source for analysis and non-analysis Cognitive model predictor variables. The SAT-M and the ACT-M are two readily available sources of Cognitive model predictor variables.

3. The Classes of 1989, 1990, and 1991 at the U.S. Air Force Academy were in transition with respect to placement and grading policies. The Class of 1989 experienced severe grade inflation in the

calculus course, which at least partially was corrected in the other two classes. Additionally, significant numbers of cadets were placed into the precalculus-calculus sequences by procedures other than those described by departmental standard operating procedures. These two conditions had a strong impact on the results of the study.

4. For the Classes of 1989, 1990, and 1991 at the U.S. Air Force Academy, the Cognitive model was at best a marginally valid placement system. The cognitive subscales were content valid but too short to be reliable. In addition, there was no empirical support for the cognitive classifications of the placement test items based on expert opinions. As was to be expected, the predictive validity of the Cognitive model increased as the number of cadets unsuccessful in calculus increased. Low predictive validity for the Class of 1989, marginal predictive validity for the Class of 1990, and acceptable predictive validity for the Class of 1991 were observed.

5. Different sets of cognitive variables were significant predictors of final calculus grades for the classes. The number of significant cognitive predictors increased with the number of unsuccessful calculus cadets. For example, only TRIGNA and MATH were significant for the Class of 1989, while all the cognitive variables were significant for the Class of 1991.

6. For the Classes of 1989, 1990, and 1991 at the U.S. Air Force Academy, the Willingham model was an acceptably valid placement system. All the Willingham model variables were significant predictors of the final calculus grades. The Willingham model yielded trends in the predictive validity similar to those found with the Cognitive model.

7. In practical terms, the various placement models displayed about the same levels of effectiveness. The Willingham, SOP, and the Computerized placement models consistently produced a small number of more correct placements, but the Cognitive model consistently provided a statistically significant better fit to final

calculus grades. The results favoring the Cognitive model are likely to be more realistic since there was a substantial margin for error in determining the number of correct placements due to the large number of placements that were unable to be classified as either correct or incorrect.

Limitations

Several limitations on the study restrict the generalizability and interpretation of the findings. These limitations are listed below.

1. The subjects in this study represented a highly select group of college freshmen. They hailed from all 52 states and survived a rigorous selection process which required both high prior academic achievement, high academic aptitude, and high military officer potential. Extreme caution should be exercised in generalizing the empirical results found in the study to other populations.

2. The extremely low reliability of the very short analysis subscales was suspected to account for the lack of evidence supporting the validity of the Cognitive model. Replication of this study using tests with larger proportions of analysis items, with corresponding increases in reliability, might produce different results.

3. The findings of the hit-and-miss tables analyses have a substantial margin for error, as the placement correctness or incorrectness for a significant portion of the hypothetically-placed cadets was unknown. These results should be interpreted cautiously given the small differences between the number of hits found for the various models.

4. The results must be considered in the context of a changing environment, since the percentage of cadets unsuccessful in calculus dramatically increased as the percentage of students enrolled in the

short and long sequences increased (no causality intended). That is to say, attempting a cross-validation in a fluid environment is not conducive to clear assessments.

Implications

The conclusions of this study imply that it is possible to develop alternative rationales for a college mathematics placement system based on various theories of learning. These systems can be logically consistent, reflect the goals and objectives of the college or university, and may be reasonably validated using historical placement data.

Colleges and universities can expect benefits from using the information about the cognitive characteristics of their placement tests. The Cognitive model precisely uses these characteristics. The main benefit is the possibility to improve predicting the success of freshmen mathematics students. Other benefits may be forthcoming from the conceptual analysis of the target mathematics sequence. Placement tests may be realigned to reflect the critical cognitive and content requirements of the sequence, which will probably increase the accuracy of predicting the success of freshmen mathematics students. In addition, inconsistencies between intended and observed sequence objectives may be identified and corrected.

Recommendations

Recommendations are provided to researchers inclined to replicate or follow-up this study. Other recommendations are included which pertain to the U. S. Air Force Academy mathematics placement system.

1. Construct the battery of placement examinations in strict accordance with the Cognitive Model of College Mathematics

Placement, beginning with the conceptual analysis of the course sequence and continuing to the cross-validation of the model. A particular focus should be to include a sufficient number of items on each cognitive subtest to produce an acceptable reliability coefficient.

2. The Cognitive model may be applicable with other sequences in mathematics and at different levels. Efforts should be made to validate the model in other settings.

3. The present model explicitly excluded use of measures of affective and social domains of behavior to predict achievement. However, the theory from which the model was developed is not so constrained; hence the model should be expanded to include affective and social variables as predictors.

4. The context of the present model was mathematics; however, the theory from which the model was developed is not limited to mathematics. Thus the usefulness of the model in other academic, business, industrial, or military areas should be investigated. This last recommendation seems most appropriate in settings where the education costs and likelihood of retraining are both high.

The findings of this study also suggest specific recommendations concerning the current placement system, the Computerized Placement Model, used by the Department of Mathematical Sciences (DFMS) at the U. S. Air Force Academy. DFMS should consider: (a) revising the algebra and trigonometry placement examinations to include more analysis items; and (b) expanding the Computerized Placement Model to include the analysis and trigonometry subscale scores as predictor variables.

APPENDICES

APPENDIX A

STANDARD OPERATING PROCEDURE (SOP) P-2

Precalculus - Calculus I Decision

IF	ACTION
ALG ≤ 14	Place in Precalculus
15 ≤ ALG ≤ 19	Scan each record for "something very positive" to indicate Calculus I placement, else place in Precalculus. Examples: (i) CALC1 ≥ 13 (ii) Math ACT ≥ 31 (iii) Math SAT ≥ 650 (iv) ACI ≥ 3300 (v) TRIG ≥ 15
20 ≤ ALG ≤ 22	Scan each record for "something negative" to indicate Precalculus placement, else place in Calculus I. Examples: (i) TRIG ≤ 11 (ii) Math ACT ≤ 26 (iii) Math SAT ≤ 600 (iv) ACI ≤ 2850
ALG ≥ 23	Essentially automatic placement into Calculus I or higher (based on CALC1 scores). Consider Prep School recommendations where appropriate. Scan for very low TRIG (i.e., TRIG ≤ 9).

APPENDIX B

CONSTRUCT AND CONTENT VALIDITY QUESTIONNAIRE

Mathematics Education
EDB 406
The University of Texas at Austin
Austin, Texas 78712

Dear Participant,

Thank you for agreeing to participate in the Placement Examination Study. Data from this questionnaire will provide information about the levels of cognitive behaviors required by the 60 items of a college placement examination for a pre-calculus course. The examination is currently used to place college freshmen into Pre-Calculus Mathematics or Differential Calculus.

Each examination item is placed on a separate page along with two questionnaire questions. **Please ignore the first question**. The second question asks you to classify the items into the four general categories of mathematical behavior as developed in the NLSMA study (Romberg & Wilson, 1969) and extended by Wilson (1971) in his model of Mathematics Achievement. The four levels of mathematics behaviors are computation, comprehension, application, and analysis.

Classification guidelines are provided on the following two pages. The guidelines consist of a definition and an example for each of the four behaviors. As you know, the classification of the items depends on the previous experience of the students. Please make your classifications assuming that the items are given to ***typical college bound high school graduates***.

Please mail the completed questionnaire back to me ***before 19 December 1988*** Do not hesitate to call me with questions concerning this questionnaire. My office phone number is **471-3747**. Finally, you are requested to protect the items on the questionnaire from release to the public as the examination is still being used.

Sincerely,

Frank J. Swehosky
Principal Investigator

REPLY TO
ATTN OF: AFIT/CIS(Major Swehosky, (512) 471-3747)

SUBJECT: Math 130 Placement Examination Study

TO: DFMS( )

1. Thank you for agreeing to participate in the study of the Math 130 Placement Examination. The questionnaire you are about to fill out will provide information about how closely the examination items corresponds to the content of the Math 130, Pre-Calculus Mathematics.

2. This packet contains the 60 items that have been used on the Algebra and Trigonometry Placement Examinations for the years 1985 through 1987. Each item is placed on a separate page along with the questionnaire questions. Please evaluate each item using the criteria provided. Do not solve the test item unless it will help you to perform the evaluation.

3. For each placement test item, the questionnaire asks two questions. The first question asks for your opinion about how well the item tests content topics on the Math 130 syllabus. A syllabus is supplied for your benefit.

4. The second question relates to the mental demands made by the items on the recent high school graduates who took the test. You are asked to classify the items into general categories of mathematical behavior. The behaviors are computation, comprehension, application, and analysis. Classification guidelines are provided on the following two pages to help you make this decision. The guidelines provide a definition and an example for each of the four behaviors.

5. The classification of the items depends on the previous experience of the student taking the examination. An item which is novel to a "typical" high school graduate may be very routine to someone with prior college mathematics experience. Please make your classification based on the assumption that the items were given to a **typical college bound high school graduate**.

6. Please return the completed packet to Lt Col Tom Curry **no later than 5 December 1988**. You may also address any questions concerning this questionnaire to Lt Col Curry or Major Frank Swehosky.

Frank J. Swehosky, Major, USAF

Guidelines for Classifying Items by Cognitive Level

1. Computation: Items designed to require straightforward manipulation of problem elements according to rules the subjects presumably have learned. Emphasis is upon performing operations, not upon deciding which operations are appropriate. (Romberg & Wilson, 1969, 39 - 40)

Example: Solve $y^2 - 3y = 18$

a. $y = 6$ or $y = -3$ b. $y = 6$ or $y = 3$
c. $y = -6$ or $y = 3$ d. $y = -6$ or $y = -3$
e. There are no real solutions

This item is a computation item because the solution demands the student to either factor the equation or to use the quadratic equation. Both of these two procedures are usually well practiced by high school graduates.

2. Comprehension: Items designed to require either recall of concepts and generalizations or transformation of problem elements from one mode to another. Emphasis is upon demonstrating understanding of concepts and their relationships, not upon using concepts to produce a solution.

Example: If $f(x) = 2x + 1$ and $g(x) = 3x - 1$, then $f(g(x)) =$:

a. $6x - 1$ b. $6x - 2$ c. $x-2$
d. $5x$ e. $6x^2 + x - 1$
(Adapted from College Entrance Examination Board, 1970, p. 54.)

This is a comprehension item because it requires an understanding of the concept of function composition. Notice that for a calculus student, this item may represent a computation item. Thus, it is important to remember your target population.

3. <u>Application:</u> Items designed to require (1) recall of relevant knowledge, (2) selection of appropriate operations, and (3) performance of the operations. Items are of a routine nature requiring the subject to use concepts in a specific context and in a way he has presumably practiced.

Example: A runner is to compete in a 26 mile marathon. What must the runner's average pace be in order to complete the race in four hours?

a. 4 mph b. 26 mph c. 6.2 mph d. 6.5 mph
e. There is not enough information to compute the average pace.

This is an application item because it is most likely a routine problem which requires the student to apply the relationship between the concepts of distance, rate, and time.

4. <u>Analysis</u>: Items designed to require a non-routine application of concepts.

Example: Find the *largest* value for x which satisfies the equation $2(8^x) + 4(8^{-x}) - 9 = 0$.

a. $\frac{-1}{3}$ b. $\frac{1}{2}$ c. $\frac{2}{3}$ d. $\frac{3}{2}$ e. 4
(Wilson et al., 1968, p. 190)

This is an analysis item because of the nonroutine nature of the problem. An understanding of exponents and equation solving are necessary. Additionally, the student must exercise a careful reading of the question in order to find the correct solution. These activities, taken together, is most likely not a practiced skill for high school graduate.

[Algebra Non-analysis Item Number 1]

Question 1.1984B

If $4^K = \frac{1}{256}$, then k equals

a. -16 b. -4 c. $-\frac{1}{4}$ d. 4

e. None of the above

1) Choose the statement below which best describes the item with regards to the content topics in the Math 130 Syllabus.

The item above ____________:

( ) does **not** test a content topic in the syllabus.
( ) is a **poor** test item of a content topic in the syllabus.
( ) is an **adequate** test item of a content topic in the syllabus.
( ) is a **good** test item of a content topic in the syllabus.
( ) is an **excellent** test item of a content topic in the syllabus..

2) Choose the category of mathematics behavior associated with this item.

( ) Computation ( ) Comprehension ( ) Application ( ) Analysis

[Algebra Analysis Item Number 15]

Question 15.1984B

If $xy = 1$, which of the following statements is <u>true</u>?

a. When $x > 1$, then $y < 0$.

b. When $x > 1$, then $y > 1$.

c. When $0 < x < 1$, then $y < 1$

d. As x increases, y decreases.

e. As x increases, y increases.

1) Choose the statement below which best describes the item with regards to the content topics in the Math 130 Syllabus.

The item above ____________:

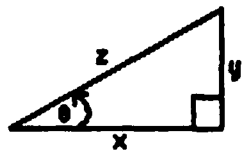
( ) does **not** test a content topic in the syllabus.
( ) is a **poor** test item of a content topic in the syllabus.
( ) is an **adequate** test item of a content topic in the syllabus.
( ) is a **good** test item of a content topic in the syllabus.
( ) is an **excellent** test item of a content topic in the syllabus..

2) Choose the category of mathematics behavior associated with this item.

( ) Computation ( ) Comprehension ( ) Application ( ) Analysis

[Trigonometry Non-analysis Item Number 1]

Question 51.1984B



Use the above figure to solve problems 51 and 52. Select answers from the following list:

a. $\frac{z}{y}$ b. $\frac{x}{y}$ c. $\frac{y}{z}$ d. $\frac{x}{z}$ e. None of the above

51. tan θ = ______

1) Choose the statement below which best describes the item with regards to the content topics in the Math 130 Syllabus.

The item above ____________:

( ) does **not** test a content topic in the syllabus.
( ) is a **poor** test item of a content topic in the syllabus.
( ) is an **adequate** test item of a content topic in the syllabus.
( ) is a **good** test item of a content topic in the syllabus.
( ) is an **excellent** test item of a content topic in the syllabus..

2) Choose the category of mathematics behavior associated with this item.

( ) Computation   ( ) Comprehension   ( ) Application   ( ) Analysis

[Trigonometry Analysis Item Number 13]

Question 63.1984B

If $\sin z = \cos z$ and if $180^\circ \leq z \leq 270^\circ$ then what is z?			
a. 1	b. 45°	c. 210°	d. 245°
e. None of the above.			

1) Choose the statement below which best describes the item with regards to the content topics in the Math 130 Syllabus.

The item above ____________:

( ) does **not** test a content topic in the syllabus.
( ) is a **poor** test item of a content topic in the syllabus.
( ) is an **adequate** test item of a content topic in the syllabus.
( ) is a **good** test item of a content topic in the syllabus.
( ) is an **excellent** test item of a content topic in the syllabus..

2) Choose the category of mathematics behavior associated with this item.

( ) Computation ( ) Comprehension ( ) Application ( ) Analysis

APPENDIX C

SUMMARY OF VALIDITY QUESTIONNAIRE RESPONSES

TABLE C.1

Validity Questionnaire[a]: Algebra Item Responses

Item	Modal Cognitive Classification	% Agreement	Modal Content Classification	% Agreement	Average Likert Score
1	2	57	4	67	3.7
2	1	86	4	67	3.7
3	2	86	4	67	4.3
4	1	71	4	67	4.3
5	2	86	NM[b]	0	4.3
6	4	43	4	67	4.0
7	2	57	4	67	4.3
8	1	86	4	67	3.7
9	2	86	4	67	4.3
10	3	86	4	67	4.3
11	4	57	4	67	3.7
12	1	100	4	67	3.3
13	1	86	4	67	4.3
14	1	71	4	67	3.7
15	4	86	4	67	2.3
16	1	100	4	67	4.3
17	1	86	4	67	4.3
18	4	86	4	67	4.3
19	2	100	5	67	4.7
20	3	86	4	100	4.0
21	1	71	4	100	4.0
22	1	100	4	67	4.3
23	1	86	4	100	4.0
24	1	71	4	67	4.3
25	2	86	5	67	4.7
26	4	43[c]	NM	0	4.0

Table C.1 (Continued)

Validity Questionnaire: Algebra Item Responses

Item	Modal Cognitive Classification	% Agreement	Modal Content Classification	% Agreement	Average Likert Score
27	2	57	4	67	4.3
28	2	43	4	100	4.0
29	2	43	4	67	4.3
30	3	86	4	67	4.3
31	2	86	4	100	4.0
32	1	100	NM	0	4.0
33	2	43	4	100	4.0
34	3	43	1	67	2.3
35	2	43	NM	0	4.0
36	2	43	4	67	4.3
37	1	71	4	100	4.0
38	1	86	3	67	3.3
39	4	57	NM	0	4.0
40	3	86	4	100	4.0
Totals					
Average		73.7		65.0	4.0
Analysis Subscale		61.9		33.3	3.7
Non-analaysis subscale		75.7		70.6	4.0

Note. [a] The construct validity results are based on seven judges and the content validity results are based on three judges. [b] No mode and no agreement between the three judges. [c] Three out of four mathematics experts judged the item as analysis; true modal classification was comprehension with 57% agreement.

Table C.2
Validity Questionnaire Results: Trigonometry Items

Item	Modal Cognitive Classification	% Agreement	Modal Content Classification	% Agreement	Average Likert Score
1	2	71	4	67	4.3
2	2	71	4	67	4.3
3	2	100	4	67	4.3
4	2	100	4	67	4.3
5	2	100	4	67	4.3
6	3	86	4	67	4.3
7	2	43	4	100	4.0
8	2	43	4	100	4.0
9	2	86	4	100	4.0
10	2	100	4	100	4.0
11	3	43	NM[a]	0	4.0
12	2	43	4	100	4.0
13	4	57	4	67	4.3
14	2	57	4	67	3.7
15	3	57	NM	0	4.0
16	2	71	4	100	4.0
17	2	71	3	67	3.3
18	2	43	4	100	4.0
19	4	29[b]	4	100	4.0
20	2	57	4	100	4.0
Total					
Average		66.4		75.0	4.1
Analysis Subscale		43.0		83.5	4.2
Non-analysis Subscale		69.0		74.1	4.1

Note. [a] No mode and no agreement between the three judges. [b] Two out of four mathematics experts judged the item as analysis; true modal classification was comprehension with 57% agreement.

APPENDIX D

STANDARD SETTING TABLES

Standards were set according to the method of Appenzellar and Kelley (1983) for each class using both the Willingham and Cognitive placement models. The following guidelines were used to suggest possible cutoff values of the Composite Score:

1. The Expected Composite Score for those persons whose course performance was just minimally satisfactory, (i.e., whose Final Calculus Grade was 2.00).

2. The Composite Score for those students with an Expected Final Calculus Grade of 2.00. This value is constant, 2.00, since all these regression equations can be shown to be:

Expected Final Calculus Grade = 1.00 Composite Score + 0.00.

3. The Composite Score for which the percentages of errors of students in each category (satisfactory, unsatisfactory) were most nearly equal.

4. The Composite Score for which the overall percentages of errors were most nearly equal.

5. The Composite Score that would have cut off, or held back, approximately the same number of students as were in the Unsastisfactory (No Pass) performance group.

6. The Composite Score that would have maximized the overall accuracy of placement.

Table 23 contains the candidate cutoff Composite Scores from applying the six guidelines to the various classes of cadets. Tables D.1 through D.21 are examples of the Kelley Tables used to support the guideline values which are in Table 23.

Table D.1
1989 Composite Score by final calculus grade: frequency distributions and descriptive statistics, using a 4.0 GPA scale for the composite scores computed with the Willingham model.

	Final Course Grades in Calculus (0-4)					
Composite Scores	0(F)	1(D)	2(C)	3(B)	4(A)	Total
3.30-4.00	0	0	0	0	0	0
3.20-3.29	0	0	2	4	3	9
3.10-3.19	0	0	6	17	19	42
3.00-3.09	0	0	18	67	30	115
2.90-2.99	0	2	39	84	26	151
2.80-2.89	2	3	53	64	19	141
2.70-2.79	1	2	24	30	22	79
2.60-2.69	0	0	11	15	5	31
2.50-2.59	0	0	3	6	3	12
2.40-2.49	1	0	0	0	1	2
0.00-2.39	0	0	0	0	0	0
Total	4	7	156	287	128	582
Mean Composite Score	2.70	2.83	2.88	2.92	2.94	2.91
Standard Deviation	0.16	0.08	0.13	0.14	0.16	0.14

Composite Score = 2.31 + 0.0088ALG + 0.032TRIG	R = 0.19
Mean Final Course Grade = 2.91	Standard Deviation = 0.77

Table D.2
1989 Composite Score by combined final calculus grade: frequency distributions and descriptive statistics, using a 4.0 GPA scale for the composite scores computed with the Willingham model.

Composite Score	Final Course Grades in Calculus: No Pass F-D	Pass C-A	Total
3.30-4.00	0	0	0
3.20-3.29	0	9	9
3.10-3.19	0	42	42
3.00-3.09	0	115	115
2.90-2.99	2	149	151
2.80-2.89	5	136	141
2.70-2.79	3	76	79
2.60-2.69	0	31	31
2.50-2.59	0	12	12
2.40-2.49	1	1	2
0.00-2.39	0	0	0
Total	11	571	582
Mean Composite Score	2.78	2.91	2.91
Standard Deviation	0.12	0.14	0.14

Composite Score = 2.31 + 0.0088ALG + 0.032TRIG	R = 0.19
Mean Final Course Grade = 2.91	Standard Deviation = 0.77

Table D.3
1989 Composite Scores using the Willingham model with possible decision scores and corresponding accuracies of placement.

	Cumulative Numbers of Students		Percent of Students in Each Placement Category			Overall Accuracy of Placement	
Placement Category	Unsat D-F (N=11)	Sat A-C (N=571)	Unsat D-F	Sat A-C	Accuracy Category	No. of Students	% of Students[1]
2.90 - up	Too High 2	315 OK	Too High 18	55 OK	Too High	2	0
		Too		Too	Correct	324	56
below 2.90	OK 9	256 Low	OK 82	45 Low	Too Low	256	44
2.80 - up	Too High 7	451 OK	Too High 64	79 OK	Too High	7	1
		Too		Too	Correct	455	78
below 2.80	OK 4	120 Low	OK 36	21 Low	Too Low	120	21
2.70 - up	Too High 10	527 OK	Too High 91	92 OK	Too High	10	2
		Too		Too	Correct	528	91
below 2.70	OK 1	44 Low	OK 9	8 Low	Too Low	44	8
2.60 - up	Too High 10	558 OK	Too High 91	98 OK	Too High	10	2
		Too		Too	Correct	559	96
below 2.60	OK 1	13 Low	OK 9	2 Low	Too Low	13	2
2.50 - up	Too High 10	570 OK	Too High 91	100 OK	Too High	10	2
		Too		Too	Correct	571	98
below 2.50	OK 1	1 Low	OK 9	0 Low	Too Low	1	0
2.40 - up	Too High 11	571 OK	Too High 100	100 OK	Too High	11	2
		Too		Too	Correct	571	98
below 2.40	OK 0	0 Low	OK 0	0 Low	Too Low	0	0
Expected Composite Score = 2.80 + .035 Final Course Grade						r = 0.19	
Final Course Grade=2.00 ⇒ Expected Composite Score=2.88						SEE^2=0.14	

Note. 1. Percentages may not sum to 100% because of rounding error.
2. Standard error of estimate.

Table D.4
1990 Composite Score by final calculus grade: frequency distributions and descriptive statistics, using a 4.0 GPA scale for the composite scores computed with the Willingham model.

	Final Course Grades in Calculus (0-4)					
Composite Scores	0(F)	1(D)	2(C)	3(B)	4(A)	Total
3.50-4.00	0	0	0	0	0	0
3.40-3.49	0	0	0	0	1	1
3.30-3.39	0	0	0	1	3	4
3.20-3.29	0	0	0	2	3	5
3.10-3.19	0	0	5	6	6	17
3.00-3.09	0	0	4	13	10	27
2.90-2.99	1	0	10	19	13	43
2.80-2.89	0	0	13	29	15	57
2.70-2.79	0	5	17	25	12	59
2.60-2.69	6	2	28	39	9	84
2.50-2.59	1	6	37	28	7	79
2.40-2.49	2	6	41	20	11	80
2.30-2.39	4	6	36	20	6	72
2.20-2.29	4	4	21	9	5	43
2.10-2.19	2	10	23	8	5	48
2.00-2.09	3	3	16	7	1	30
1.90-1.99	1	1	5	5	2	14
1.80-1.89	1	2	5	5	1	14
1.70-1.79	0	0	3	3	0	6
1.60-1.69	0	3	4	1	0	8
1.50-1.59	0	0	0	0	0	0
1.40-1.49	0	0	1	0	0	1
0.00-1.39	0	0	0	0	0	0
Total	25	48	269	240	110	692
Mean Composite Score	2.36	2.29	2.44	2.59	2.72	2.52
Standard Deviation	0.27	0.30	0.31	0.32	0.34	0.34
Composite Score = 1.18 + 0.044ALG + 0.027TRIG						R = 0.35
Mean Final Course Grade = 2.52				Standard Deviation = 0.96		

Table D.5

1990 Composite Score by combined final calculus grade: frequency distributions and descriptive statistics, using a 4.0 GPA scale for the composite scores computed with the Willingham model.

Composite Score	Final Course Grades in Calculus: No Pass F-D	Pass C-A	Total
3.50-4.00	0	0	0
3.40-3.49	0	1	1
3.30-3.39	0	4	4
3.20-3.29	0	5	5
3.10-3.19	0	17	17
3.00-3.09	0	27	27
2.90-2.99	1	42	43
2.80-2.89	0	57	57
2.70-2.79	5	54	59
2.60-2.69	8	76	84
2.50-2.59	7	72	79
2.40-2.49	8	72	80
2.30-2.39	10	62	72
2.20-2.29	8	35	43
2.10-2.19	12	36	48
2.00-2.09	6	24	30
1.90-1.99	2	12	14
1.80-1.89	3	11	14
1.70-1.79	0	6	6
1.60-1.69	3	5	8
1.50-1.59	0	0	0
1.40-1.49	0	1	1
0.00-1.39	0	0	0
Total	73	619	695
Mean Composite Score	2.31	2.55	2.52
Standard Deviation	0.29	0.34	0.34

Composite Score = 1.18 + 0.04[illegible]LG + 0.027TRIG	R = 0.35
Mean Final Course Grade = 2.52	Standard Deviation = 0.96

Table D.6
1990 Composite Scores using the Willingham model with possible decision scores and corresponding accuracies of placement.

	Cumulative Numbers of Students		Percent of Students in Each Placement Category			Overall Accuracy of Placement	
Placement Category	Unsat D-F (N=11)	Sat A-C (N=571)	Unsat D-F	Sat A-C	Accuracy Category	No. of Students	% of Students[1]
2.40- up	Too High 29	427 OK	Too High 40	69 OK	Too High	29	4
					Correct	471	68
below 2.40	OK 44	192 Too Low	OK 60	31 Too Low	Too Low	192	28
2.30- up	Too High 39	489 OK	Too High 53	79 OK	Too High	39	6
					Correct	523	76
below 2.30	OK 34	130 Too Low	OK 47	21 Too Low	Too Low	130	19
2.20- up	Too High 47	524 OK	Too High 64	85 OK	Too High	47	7
					Correct	550	79
below 2.20	OK 26	95 Too Low	OK 36	15 Too Low	Too Low	95	14
2.10 - up	Too High 59	560 OK	Too High 81	90 OK	Too High	59	9
					Correct	574	83
below 2.10	OK 14	59 Too Low	OK 19	10 Too Low	Too Low	59	9
2.00- up	Too High 65	584 OK	Too High 89	94 OK	Too High	65	9
					Correct	592	86
below 2.00	OK 8	35 Too Low	OK 11	6 Too Low	Too Low	35	5
1.90 - up	Too High 67	596 OK	Too High 92	96 OK	Too High	67	10
					Correct	602	87
below 1.90	OK 6	23 Too Low	OK 8	4 Too Low	Too Low	23	3
1.80 - up	Too High 70	607 OK	Too High 96	98 OK	Too High	70	10
					Correct	610	88
below 1.80	OK 3	12 Too Low	OK 4	2 Too Low	Too Low	12	2
1.70 - up	Too High 70	613 OK	Too High 96	99 OK	Too High	70	10
					Correct	616	89
below 1.70	OK 3	6 Too Low	OK 4	1 Too Low	Too Low	6	1
1.60 - up	Too High 73	618 OK	Too High 100	100 OK	Too High	73	11
					Correct	618	89
below 1.60	OK 0	1 Too Low	OK 0	0 Too Low	Too Low	1	0
1.50 - up	Too High 73	618 OK	Too High 100	100 OK	Too High	73	11
					Correct	618	89
below 1.50	OK 0	1 Too Low	OK 0	0 Too Low	Too Low	1	0
1.40 - up	Too High 73	619 OK	Too High 100	100 OK	Too High	73	11
					Correct	619	89
below 1.40	OK 0	0 Too Low	OK 0	0 Too Low	Too Low	0	0
Expected Composite Score = 2.21 + 0.12 Final Course Grade						r = 0.35	
Final Course Grade = 2.00 ⇒ Expected Composite Score = 2.46						SEE[2] = 0.32	

Note. 1. Percentages may not sum to 100% because of rounding error. 2. Standard error of estimate.

Table D.7
1991 Composite Score by final calculus grade: frequency distributions and descriptive statistics, using a 4.0 GPA scale for the composite scores computed with the Willingham model.

Composite Scores	0(F)	1(D)	2(C)	3(B)	4(A)	Total
	Final Course Grades in Calculus (0-4)					
3.50-4.00	0	0	0	1	6	7
3.40-3.49	0	0	1	3	11	15
3.30-3.39	0	0	1	8	7	16
3.20-3.29	1	0	4	5	4	14
3.10-3.19	0	0	8	13	18	39
3.00-3.09	0	0	3	17	9	29
2.90-2.99	2	0	12	22	9	45
2.80-2.89	1	0	12	15	14	42
2.70-2.79	0	1	20	18	8	47
2.60-2.69	1	2	16	24	5	48
2.50-2.59	3	3	25	32	4	67
2.40-2.49	2	8	26	12	5	53
2.30-2.39	4	2	24	16	2	48
2.20-2.29	4	6	24	12	6	52
2.10-2.19	1	2	24	13	3	43
2.00-2.09	1	7	13	7	9	37
1.90-1.99	6	5	12	6	2	31
1.80-1.89	4	3	11	5	4	27
1.70-1.79	4	4	17	4	1	30
1.60-1.69	3	2	6	1	1	13
1.50-1.59	4	0	6	2	2	14
1.40-1.49	2	2	2	2	0	8
1.30-1.39	4	3	2	0	0	9
1.20-1.29	1	1	1	0	0	3
0.00-1.19	2	2	1	0	0	5
Total	50	53	271	238	130	742
Mean Composite Score	1.97	2.03	2.35	2.62	2.80	2.47
Standard Deviation	0.50	0.43	0.44	0.42	0.52	0.52

Composite Score = 0.36 + 0.055ALG + 0.067TRIG	R = 0.48
Mean Final Course Grade = 2.47	Standard Deviation = 1.07

Table D.8
1991 Composite Score by combined final calculus grade: frequency distributions and descriptive statistics, using a 4.0 GPA scale for the composite scores computed with the Willingham model.

Composite Score	Final Course Grades in Calculus: No Pass F-D	Pass C-A	Total
3.50-4.00	0	7	7
3.40-3.49	0	15	15
3.30-3.39	0	16	16
3.20-3.29	1	13	14
3.10-3.19	0	39	39
3.00-3.09	0	29	29
2.90-2.99	2	43	45
2.80-2.89	1	41	42
2.70-2.79	1	46	47
2.60-2.69	3	45	48
2.50-2.59	6	61	67
2.40-2.49	10	43	53
2.30-2.39	6	42	48
2.20-2.29	10	42	52
2.10-2.19	3	40	43
2.00-2.09	8	29	37
1.90-1.99	11	20	31
1.80-1.89	7	20	27
1.70-1.79	8	22	30
1.60-1.69	5	8	13
1.50-1.59	4	10	14
1.40-1.49	4	4	8
1.30-1.39	7	2	9
1.20-1.29	2	1	3
0.00-1.19	4	1	5
Total	103	639	742
Mean Composite Score	2.00	2.54	2.47
Standard Deviation	0.46	0.48	0.52
Composite Score = 0.36 + 0.055ALG + 0.067TRIG			R = 0.48
Mean Final Course Grade = 2.47		Standard Deviation = 1.07	

Table D.9
1991 Composite Scores using the Willingham model with possible decision scores and corresponding accuracies of placement.

	Cumulative Numbers of Students		Percent of Students in Each Placement Category			Overall Accuracy of Placement	
Placement Category	Unsat D-F (N=11)	Sat A-C (N=571)	Unsat D-F	Sat A-C	Accuracy Category	No. of Students	% of Students[1]
2.40 - up	Too High 24	398 OK	Too High 23	62 OK	Too High	24	3
		Too		Too	Correct	477	64
below 2.40	OK 79	241 Low	OK 77	38 Low	Too Low	241	32
2.30- up	Too High 30	440 OK	Too High 29	69 OK	Too High	30	4
		Too		Too	Correct	513	69
below 2.30	OK 73	199 Low	OK 71	31 Low	Too Low	199	27
2.20- up	Too High 40	482 OK	Too High 39	75 OK	Too High	40	5
		Too		Too	Correct	545	73
below 2.20	OK 63	157 Low	OK 61	25 Low	Too Low	157	21
2.10 - up	Too High 43	522 OK	Too High 42	82 OK	Too High	43	6
		Too		Too	Correct	582	78
below 2.10	OK 60	117 Low	OK 58	18 Low	Too Low	117	16
2.00- up	Too High 51	551 OK	Too High 50	86 OK	Too High	51	7
		Too		Too	Correct	603	81
below 2.00	OK 52	88 Low	OK 50	14 Low	Too Low	88	12
1.90 - up	Too High 62	571 OK	Too High 60	89 OK	Too High	62	8
		Too		Too	Correct	612	82
below 1.90	OK 41	68 Low	OK 40	11 Low	Too Low	68	9
1.80 - up	Too High 69	591 OK	Too High 67	92 OK	Too High	69	9
		Too		Too	Correct	625	84
below 1.80	OK 34	48 Low	OK 33	8 Low	Too Low	48	6
1.70 - up	Too High 77	613 OK	Too High 75	96 OK	Too High	77	10
		Too		Too	Correct	639	86
below 1.70	OK 26	26 Low	OK 25	4 Low	Too Low	26	4
1.60 - up	Too High 82	621 OK	Too High 80	97 OK	Too High	82	11
		Too		Too	Correct	642	87
below 1.60	OK 21	18 Low	OK 20	3 Low	Too Low	18	2
1.50 - up	Too High 86	631 OK	Too High 83	99 OK	Too High	86	12
		Too		Too	Correct	648	87
below 1.50	OK 17	8 Low	OK 17	1 Low	Too Low	8	1
1.40 - up	Too High 90	635 OK	Too High 87	99 OK	Too High	90	12
		Too		Too	Correct	648	87
below 1.40	OK 13	4 Low	OK 13	1 Low	Too Low	4	1
Expected Composite Score = 1.89 + 0.23 Final Course Grade						r = 0.48	
Final Course Grade = 2.00 ⇒ Expected Composite Score = 2.36						SEE^2 = 0.45	

Note. 1. Percentages may not sum to 100% because of rounding error. 2. Standard error of estimate.

Table D.10
1989 Composite Score by final calculus grade: frequency distributions and descriptive statistics, using a 4.0 GPA scale for the composite scores computed with the Cognitive model.

	Final Course Grades in Calculus (0-4)					
Composite Scores	0(F)	1(D)	2(C)	3(B)	4(A)	Total
3.50-4.00	0	0	0	0	0	0
3.40-3.49	0	0	0	1	0	1
3.30-3.39	0	0	2	5	1	8
3.20-3.29	0	0	0	7	16	23
3.10-3.19	0	0	13	26	21	60
3.00-3.09	0	1	18	60	23	102
2.90-2.99	0	2	28	61	13	104
2.80-2.89	1	1	33	59	21	115
2.70-2.79	0	0	36	39	23	98
2.60-2.69	1	2	15	20	9	47
2.50-2.59	1	1	9	8	0	19
2.40-2.49	1	0	2	1	1	5
0.00-2.39	0	0	0	0	0	0
Total	4	7	156	287	128	582
Mean Composite Score	2.64	2.81	2.85	2.92	2.96	2.91
Standard Deviation	0.17	0.20	0.17	0.17	0.20	0.18

Composite Score = -1.04 - .033ALGB + .0057ALGBNA + .059TRIGA + .027TRIGNA + .0018MATH

Mean Final Course Grade = 2.91	Standard Deviation = 0.77	R = 0.24

Table D.11
1989 Composite Score by combined final calculus grade: frequency distributions and descriptive statistics, using a 4.0 GPA scale for the composite scores computed with the Cognitive model.

Composite Score	Final Course Grades in Calculus: No Pass F-D	Pass C-A	Total
3.50-4.00	0	0	0
3.40-3.49	0	1	1
3.30-3.39	0	8	8
3.20-3.29	0	23	23
3.10-3.19	0	60	60
3.00-3.09	1	101	102
2.90-2.99	2	102	104
2.80-2.89	2	113	115
2.70-2.79	0	98	98
2.60-2.69	3	44	47
2.50-2.59	2	17	19
2.40-2.49	1	4	5
0.00-2.39	0	0	0
Total	11	571	582
Mean Composite Score	2.75	2.91	2.91
Standard Deviation	0.20	0.18	0.18

Composite Score = -1.04 - .033ALGB + .0057ALGBNA + .059TRIGA + .027TRIGNA + .0018MATH

Mean Final Course Grade = 2.91	Standard Deviation = 0.77	R = 0.24

Table D.12
1989 Composite Scores using the Cognitive model with possible decision scores and corresponding accuracies of placement.

	Cumulative Numbers of Students		Percent of Students in Each Placement Category			Overall Accuracy of Placement	
Placement Category	Unsat D-F (N=11)	Sat A-C (N=571)	Unsat D-F	Sat A-C	Accuracy Category	No. of Students	% of Students[1]
3.00 - up	Too High 1	193 OK	Too High 9	34 OK	Too High	1	0
		Too		Too	Correct	203	35
below 3.00	OK 10	378 Low	OK 91	66 Low	Too Low	378	65
2.90 - up	Too High 3	295 OK	Too High 27	52 OK	Too High	3	1
		Too		Too	Correct	303	52
below 2.90	OK 8	276 Low	OK 73	48 Low	Too Low	276	47
2.80 - up	Too High 5	408 OK	Too High 45	71 OK	Too High	5	1
		Too		Too	Correct	414	71
below 2.80	OK 6	163 Low	OK 55	29 Low	Too Low	163	28
2.70 - up	Too High 5	506 OK	Too High 45	89 OK	Too High	5	1
		Too		Too	Correct	512	88
below 2.70	OK 6	65 Low	OK 55	11 Low	Too Low	65	11
2.60 - up	Too High 8	550 OK	Too High 73	96 OK	Too High	8	1
		Too		Too	Correct	553	95
below 2.60	OK 3	21 Low	OK 27	4 Low	Too Low	21	4
2.50 - up	Too High 10	567 OK	Too High 91	99 OK	Too High	10	2
		Too		Too	Correct	568	98
below 2.50	OK 1	4 Low	OK 9	1 Low	Too Low	4	1
2.40 - up	Too High 11	571 OK	Too High 100	100 OK	Too High	11	2
		Too		Too	Correct	571	98
below 2.40	OK 0	0 Low	OK 0	0 Low	Too Low	0	0
Expected Composite Score = 2.74 + .057 Final Course Grade						r = 0.24	
Final Course Grade = 2.00 ➡ Expected Composite Score = 2.86						SEE^2 = 0.18	

Note. 1. Percentages may not sum to 100% because of rounding error. 2. Standard error of estimate.

Table D.13
1990 Composite Score by final calculus grade: frequency distributions and descriptive statistics, using a 4.0 GPA scale for the composite scores computed with the Cognitive model.

	Final Course Grades in Calculus (0-4)					
Composite Scores	0(F)	1(D)	2(C)	3(B)	4(A)	Total
3.50-4.00	0	0	0	0	2	2
3.40-3.49	0	0	0	1	1	2
3.30-3.39	0	0	0	5	2	7
3.20-3.29	0	0	2	1	3	6
3.10-3.19	0	0	3	5	8	16
3.00-3.09	0	0	7	12	11	30
2.90-2.99	0	0	2	21	13	36
2.80-2.89	1	1	20	37	12	71
2.70-2.79	3	2	16	30	16	67
2.60-2.69	1	3	21	21	8	54
2.50-2.59	1	8	38	24	6	77
2.40-2.49	3	7	38	21	5	74
2.30-2.39	6	1	35	21	8	71
2.20-2.29	4	8	28	10	6	56
2.10-2.19	2	6	19	9	3	39
2.00-2.09	2	3	12	8	4	29
1.90-1.99	0	3	16	5	2	26
1.80-1.89	1	3	4	7	0	15
1.70-1.79	0	0	3	1	0	4
1.60-1.69	1	2	2	0	0	5
1.50-1.59	0	1	3	0	0	4
1.40-1.49	0	0	0	1	0	1
0.00-1.39	0	0	0	0	0	0
Total	25	48	273	240	111	692
Mean Composite Score	2.34	2.28	2.42	2.61	2.74	2.52
Standard Deviation	0.28	0.30	0.32	0.34	0.36	0.36

Composite Score = .16 -.00012ALGB +.041ALGBNA +.018TRIGA +.033TRIGNA +.0019MATH		
Mean Final Course Grade = 2.52	Standard Deviation = 0.96	R = 0.37

Table D.14
1990 Composite Score by combined final calculus grade: frequency distributions and descriptive statistics, using a 4.0 GPA scale for the composite scores computed with the Cognitive model.

Composite Score	Final Course Grades in Calculus: No Pass F-D	Pass C-A	Total
3.50-4.00	0	2	2
3.40-3.49	0	2	2
3.30-3.39	0	7	7
3.20-3.29	0	6	6
3.10-3.19	0	16	16
3.00-3.09	0	30	30
2.90-2.99	0	36	36
2.80-2.89	2	69	71
2.70-2.79	5	62	67
2.60-2.69	4	50	54
2.50-2.59	9	68	77
2.40-2.49	10	64	74
2.30-2.39	7	64	71
2.20-2.29	12	44	56
2.10-2.19	8	31	39
2.00-2.09	5	24	29
1.90-1.99	3	23	26
1.80-1.89	4	11	15
1.70-1.79	0	4	4
1.60-1.69	3	2	5
1.50-1.59	1	3	4
1.40-1.49	0	1	1
0.00-1.39	0	0	0
Total	73	619	692
Mean Composite Score	2.30	2.55	2.52
Standard Deviation	0.30	0.36	0.36

Composite Score = .16 -.00012ALGB +.041ALGBNA +.018TRIGA +.033TRIGNA +.0019MATH		
Mean Final Course Grade = 2.52	Standard Deviation = 0.96	R = 0.37

Table D.15
1990 Composite Scores using the Cognitive model with possible decision scores and corresponding accuracies of placement.

	Cumulative Numbers of Students		Percent of Students in Each Placement Category			Overall Accuracy of Placement	
Placement Category	Unsat D-F (N=11)	Sat A-C (N=571)	Unsat D-F	Sat A-C	Accuracy Category	No. of Students	% of Students[1]
2.50- up	Too High 20	348 OK	Too High 27	56 OK	Too High	20	3
		Too		Too	Correct	401	58
below 2.50	OK 53	271 Low	OK 73	44 Low	Too Low	271	39
2.40- up	Too High 30	412 OK	Too High 41	67 OK	Too High	30	4
		Too		Too	Correct	455	66
below 2.40	OK 43	207 Low	OK 59	33 Low	Too Low	207	30
2.30- up	Too High 37	476 OK	Too High 51	77 OK	Too High	37	5
		Too		Too	Correct	512	74
below 2.30	OK 36	143 Low	OK 49	23 Low	Too Low	143	21
2.20- up	Too High 49	520 OK	Too High 67	84 OK	Too High	49	7
		Too		Too	Correct	544	79
below 2.20	OK 24	99 Low	OK 33	16 Low	Too Low	99	14
2.10 - up	Too High 57	551 OK	Too High 78	89 OK	Too High	57	8
		Too		Too	Correct	567	82
below 2.10	OK 16	68 Low	OK 22	11 Low	Too Low	68	10
2.00- up	Too High 62	575 OK	Too High 85	93 OK	Too High	62	9
		Too		Too	Correct	586	85
below 2.00	OK 11	44 Low	OK 15	7 Low	Too Low	44	6
1.90 - up	Too High 65	598 OK	Too High 89	97 OK	Too High	65	9
		Too		Too	Correct	606	88
below 1.90	OK 8	21 Low	OK 11	3 Low	Too Low	21	3
1.80 - up	Too High 69	609 OK	Too High 95	98 OK	Too High	69	10
		Too		Too	Correct	613	89
below 1.80	OK 4	10 Low	OK 5	2 Low	Too Low	10	1
1.70 - up	Too High 69	613 OK	Too High 95	99 OK	Too High	69	10
		Too		Too	Correct	617	89
below 1.70	OK 4	6 Low	OK 5	1 Low	Too Low	6	1
1.60 - up	Too High 72	615 OK	Too High 99	99 OK	Too High	72	10
		Too		Too	Correct	616	89
below 1.60	OK 1	4 Low	OK 1	1 Low	Too Low	4	1
1.50 - up	Too High 73	618 OK	Too High 100	100 OK	Too High	73	11
		Too		Too	Correct	618	89
below 1.50	OK 0	1 Low	OK 0	0 Low	Too Low	1	0
Expected Composite Score = 2.17 + 0.14 Final Course Grade						r = 0.37	
Final Course Grade = 2.00 ⇒ Expected Composite Score = 2.45						SEE[2] = 0.33	

Note. 1. Percentages may not sum to 100% because of rounding error. 2. Standard error of estimate.

Table D.16
1991 Composite Score by final calculus grade: frequency distributions and descriptive statistics, using a 4.0 GPA scale for the composite scores computed with the Cognitive model.

	Final Course Grades in Calculus (0-4)					
Composite Scores	0(F)	1(D)	2(C)	3(B)	4(A)	Total
3.50-4.00	0	0	1	2	10	13
3.40-3.49	0	0	1	3	13	17
3.30-3.39	1	0	2	9	4	16
3.20-3.29	0	0	4	10	16	30
3.10-3.19	0	0	6	11	9	26
3.00-3.09	0	0	7	16	6	29
2.90-2.99	1	0	12	11	4	28
2.80-2.89	0	0	8	22	10	40
2.70-2.79	0	2	21	24	11	58
2.60-2.69	2	1	18	21	5	47
2.50-2.59	1	3	26	17	6	53
2.40-2.49	1	3	24	18	7	53
2.30-2.39	5	4	22	11	4	46
2.20-2.29	4	8	19	20	5	56
2.10-2.19	2	4	20	16	5	47
2.00-2.09	6	5	17	10	2	40
1.90-1.99	4	5	16	5	8	38
1.80-1.89	3	3	13	3	1	23
1.70-1.79	4	3	11	7	3	28
1.60-1.69	4	4	6	0	0	14
1.50-1.59	2	1	8	1	1	13
1.40-1.49	1	2	5	1	0	9
1.30-1.39	1	0	2	0	0	3
1.20-1.29	4	2	0	0	0	6
0.00-1.19	4	3	2	0	0	9
Total	50	53	271	238	130	742
Mean Composite Score	1.92	2.00	2.34	2.62	2.83	2.46
Standard Deviation	0.50	0.44	0.46	0.43	0.52	0.53

Composite Score = -.27 +.13ALGB +.036ALGBNA +.16TRIGA +.055TRIGNA +.0012MATH		
R = 0.50	SEE = 0.93	Mean Final Course Grade = 2.47 Standard Deviation = 1.07

Table D.17
1991 Composite Score by combined final calculus grade: frequency distributions and descriptive statistics, using a 4.0 GPA scale for the composite scores computed with the Cognitive model.

Composite Score	Final Course Grades in Calculus: No Pass F-D	Pass C-A	Total
3.50-4.00	0	13	13
3.40-3.49	0	17	17
3.30-3.39	1	15	16
3.20-3.29	0	30	30
3.10-3.19	0	26	26
3.00-3.09	0	29	29
2.90-2.99	1	27	28
2.80-2.89	0	40	40
2.70-2.79	2	56	58
2.60-2.69	3	44	47
2.50-2.59	4	49	53
2.40-2.49	4	49	53
2.30-2.39	9	37	46
2.20-2.29	12	44	56
2.10-2.19	6	41	47
2.00-2.09	11	29	40
1.90-1.99	9	29	38
1.80-1.89	6	17	23
1.70-1.79	7	21	28
1.60-1.69	8	6	14
1.50-1.59	3	10	13
1.40-1.49	3	6	9
1.30-1.39	1	2	3
1.20-1.29	6	0	6
0.00-1.19	7	2	9
Total	103	639	742
Mean Composite Score	1.96	2.55	2.46
Standard Deviation	0.47	0.50	0.53

Composite Score = -.27 +.13ALGB +.036ALGBNA +.16TRIGA +.055TRIGNA +.0012MATH

R = 0.50 | SEE = 0.93 | Mean Final Course Grade = 2.47 Standard Deviation = 1.07

Table D.18
1991 Composite Scores using the Cognitive model with possible decision scores and corresponding accuracies of placement.

	Cumulative Numbers of Students		Percent of Students in Each Placement Category			Overall Accuracy of Placement	
Placement Category	Unsat D-F (N=11)	Sat A-C (N=571)	Unsat D-F	Sat A-C	Accuracy Category	No. of Students	% of Students[1]
2.30- up	Too High 24	432 OK	Too High 23	68 OK	Too High	24	3
		Too		Too	Correct	511	69
below 2.30	OK 79	207 Low	OK 77	32 Low	Too Low	207	28
2.20- up	Too High 36	476 OK	Too High 35	74 OK	Too High	36	5
		Too		Too	Correct	543	73
below 2.20	OK 67	163 Low	OK 65	26 Low	Too Low	163	22
2.10 - up	Too High 42	517 OK	Too High 41	81 OK	Too High	42	6
		Too		Too	Correct	578	78
below 2.10	OK 61	122 Low	OK 59	19 Low	Too Low	122	16
2.00- up	Too High 53	546 OK	Too High 51	85 OK	Too High	53	7
		Too		Too	Correct	596	80
below 2.00	OK 50	93 Low	OK 49	15 Low	Too Low	93	13
1.90 - up	Too High 62	575 OK	Too High 60	90 OK	Too High	62	8
		Too		Too	Correct	616	83
below 1.90	OK 41	64 Low	OK 40	10 Low	Too Low	64	9
1.80 - up	Too High 68	592 OK	Too High 66	93 OK	Too High	68	9
		Too		Too	Correct	627	85
below 1.80	OK 35	47 Low	OK 34	7 Low	Too Low	47	6
1.70 - up	Too High 75	613 OK	Too High 73	96 OK	Too High	75	10
		Too		Too	Correct	641	86
below 1.70	OK 28	26 Low	OK 27	4 Low	Too Low	26	4
1.60 - up	Too High 83	619 OK	Too High 81	97 OK	Too High	83	11
		Too		Too	Correct	639	86
below 1.60	OK 20	20 Low	OK 19	3 Low	Too Low	20	3
1.50 - up	Too High 86	629 OK	Too High 83	98 OK	Too High	86	12
		Too		Too	Correct	646	87
below 1.50	OK 17	10 Low	OK 17	2 Low	Too Low	10	1
1.40 - up	Too High 89	635 OK	Too High 86	99 OK	Too High	89	12
		Too		Too	Correct	649	87
below 1.40	OK 14	4 Low	OK 14	1 Low	Too Low	4	1
1.30 - up	Too High 90	637 OK	Too High 87	100 OK	Too High	90	12
		Too		Too	Correct	650	88
below 1.30	OK 13	2 Low	OK 13	0 Low	Too Low	2	0
Expected Composite Score = 1.85 + 0.25 Final Course Grade						r = 0.50	
Final Course Grade = 2.00 ⇒ Expected Composite Score = 2.35						SEE^2 = 0.46	

Note. 1. Percentages may not sum to 100% because of rounding error. 2. Standard error of estimate.

Table D.19

1991 (reduced) Composite Score by final calculus grade: frequency distributions and descriptive statistics, using a 4.0 GPA scale for the composite scores computed with the Cognitive model.

	Final Course Grades in Calculus (0-4)					
Composite Scores	0(F)	1(D)	2(C)	3(B)	4(A)	Total
3.50-4.00	0	0	0	1	2	3
3.40-3.49	0	0	1	1	7	9
3.30-3.39	0	0	2	4	8	14
3.20-3.29	1	0	1	9	4	15
3.10-3.19	0	0	4	8	11	23
3.00-3.09	0	0	8	11	5	24
2.90-2.99	0	0	8	12	3	23
2.80-2.89	1	0	11	8	6	26
2.70-2.79	0	0	9	20	6	35
2.60-2.69	0	3	18	18	10	49
2.50-2.59	2	2	24	25	5	58
2.40-2.49	2	2	16	8	5	33
2.30-2.39	1	6	23	12	4	46
2.20-2.29	5	3	20	16	6	50
2.10-2.19	3	6	15	11	1	36
2.00-2.09	4	4	21	12	3	[illegible]4
1.90-1.99	4	6	10	4	6	30
1.80-1.89	5	3	16	5	4	33
1.70-1.79	2	2	7	5	1	17
1.60-1.69	4	5	9	0	1	19
1.50-1.59	1	1	6	1	0	9
1.40-1.49	1	1	5	1	0	8
1.30-1.39	1	1	3	0	0	5
1.20-1.29	4	2	1	0	0	7
0.00-1.19	4	3	1	0	0	8
Total	45	50	239	192	98	624
Mean Composite Score	1.90	1.97	2.30	2.57	2.74	2.40
Standard Deviation	0..48	0.42	0.45	0.42	0.51	0.52

Composite Score = -.18 + .13 ALGB + .031ALGBNA + .18 TRIGA + .044 TRIGNA + .0013MATH

Mean Final Course Grade = 2.40	Standard Deviation = 1.07	R = 0.48

Table D.20
1991 (reduced) Composite Score by combined final calculus grade: frequency distributions and descriptive statistics, using a 4.0 GPA scale for the composite scores computed with the Cognitive model.

Composite Score	Final Course Grades in Calculus: No Pass F-D	Pass C-A	Total
3.50-4.00	0	3	3
3.40-3.49	0	9	9
3.30-3.39	0	14	14
3.20-3.29	1	14	15
3.10-3.19	0	23	23
3.00-3.09	0	24	24
2.90-2.99	0	23	23
2.80-2.89	1	25	26
2.70-2.79	0	35	35
2.60-2.69	3	46	49
2.50-2.59	4	54	58
2.40-2.49	4	29	33
2.30-2.39	7	39	46
2.20-2.29	8	42	50
2.10-2.19	9	27	36
2.00-2.09	8	36	44
1.90-1.99	10	20	30
1.80-1.89	8	25	33
1.70-1.79	4	13	17
1.60-1.69	9	10	19
1.50-1.59	2	7	9
1.40-1.49	2	6	8
1.30-1.39	2	3	5
1.20-1.29	6	1	7
0.00-1.19	7	1	8
Total	95	529	624
Mean Composite Score	1.94	2.48	2.40
Standard Deviation	0.45	0.48	0.52

Composite Score = -.18 + .13ALGB + .031ALGBNA + .18TRIGA + .044TRIGNA + .0013MATH

Mean Final Course Grade = 2.40	Standard Deviation = 1.07	R = 0.48

Table D.21
1991 (reduced) Composite Scores using the Cognitive model with possible decision scores and corresponding accuracies of placement For comparing Cognitive vs. Computerized models.

	Cumulative Numbers of Students		Percent of Students in Each Placement Category			Overall Accuracy of Placement	
Placement Category	Unsat D-F (N=11)	Sat A-C (N=571)	Unsat D-F	Sat A-C	Accuracy Category	No. of Students	% of Students[1]
2.30- up	Too High 20	338 OK	Too High 21	64 OK	Too High	20	3
		Too		Too	Correct	413	66
below 2.30	OK 75	191 Low	OK 79	36 Low	Too Low	191	31
2.20- up	Too High 28	380 OK	Too High 29	72 OK	Too High	28	4
		Too		Too	Correct	447	72
below 2.20	OK 67	149 Low	OK 71	28 Low	Too Low	149	24
2.10 - up	Too High 37	407 OK	Too High 39	77 OK	Too High	37	6
		Too		Too	Correct	465	75
below 2.10	OK 58	122 Low	OK 61	23 Low	Too Low	122	20
2.00- up	Too High 45	443 OK	Too High 47	84 OK	Too High	45	7
		Too		Too	Correct	493	79
below 2.00	OK 50	86 Low	OK 53	16 Low	Too Low	86	14
1.90 - up	Too High 55	463 OK	Too High 58	88 OK	Too High	55	9
		Too		Too	Correct	503	81
below 1.90	OK 40	66 Low	OK 42	12 Low	Too Low	66	11
1.80 - up	Too High 63	488 OK	Too High 66	92 OK	Too High	63	10
		Too		Too	Correct	520	83
below 1.80	OK 32	41 Low	OK 34	8 Low	Too Low	41	7
1.70 - up	Too High 67	501 OK	Too High 71	95 OK	Too High	67	11
		Too		Too	Correct	529	85
below 1.70	OK 28	28 Low	OK 29	5 Low	Too Low	28	4
1.60 - up	Too High 76	511 OK	Too High 80	97 OK	Too High	76	12
		Too		Too	Correct	530	85
below 1.60	OK 19	18 Low	OK 20	3 Low	Too Low	18	3
1.50 - up	Too High 78	518 OK	Too High 82	98 OK	Too High	78	13
		Too		Too	Correct	535	86
below 1.50	OK 17	11 Low	OK 18	2 Low	Too Low	11	2
1.40 - up	Too High 80	524 OK	Too High 84	99 OK	Too High	80	13
		Too		Too	Correct	539	86
below 1.40	OK 15	5 Low	OK 16	1 Low	Too Low	5	1
1.30 - up	Too High 82	527 OK	Too High 86	100 OK	Too High	82	13
		Too		Too	Correct	540	87
below 1.30	OK 13	2 Low	OK 14	0 Low	Too Low	2	0
Expected Composite Score = 1.84 + 0.23 Final Course Grade						r = 0.48	
Final Course Grade = 2.00 ⇒ Expected Composite Score = 2.31						$SEE^2 = 0.45$	

Note. 1. Percentages may not sum to 100% because of rounding error. 2. Standard error of estimate.

REFERENCES

Ahrens, S. (1980). Analysis and classification of entering freshmen mathematics students using multiple discriminant function analysis. Paper presented at the Annual Meeting of the Southern Association of Institutional Research, Louisville, KY, Fall, 1980. (ERIC Document Reproduction Service No. 205 146)

Aleamoni, L. M. (1979). Methods of implementing college placement and exemption programs. New York: College Entrance Examination Board.

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), Educational measurement. Washington D. C.: American Council on Education.

Angoff, W. H. & Dyer, H. S. (1971). The admissions testing program. In W. H. Angoff (Ed.) The College Board admissions testing program: A technical report on research and development activities relating to the Scholastic Aptitude Test and Achievement Tests. New York: College Entrance Examination Board.

Appenzellar, A. B. & Kelley, H. P. (1983). Validity study of CLEP subject examination in calculus with elementary functions for use in credit by examination at U.T. Austin, Spring and Fall 1982 (Research Bulletin RB-83-3). Austin: University of Texas at Austin, Measurement and Evaluation Center.

Avital, S. M. & Shettleworth, S. J. (1968). Objectives for mathematics learning: Some ideas for the teacher (Bulletin No. 3). Ontario, Canada: The Ontario Institute for Studies in Education.

Bingham, R. L. (1972). An investigation into the relationship between advanced placement in mathematics and performance in first semester calculus at the University of Texas at Austin. (Doctoral dissertation, University of Texas at Austin, 1972). Dissertation Abstracts International, 33, 4865A.

Boudot, J. R. (1987). Automated placement of students into mathematics curricula. Unpublished paper, United States Air Force Academy, Department of Mathematical Sciences, Colorado.

Bridgeman, B. (1980). Validation of the descriptive tests of mathematics. Princeton, N.J.: Educational Testing Service. (ERIC Document Reproduction Service No. ED 196 949).

Burr, I. W. (1974). Applied statistical methods. New York, NY: Academic Press.

Cederberg, J. & Harvey, J. G. (1987, Spring/Summer). Questionnaire results. Placement Test Newsletter, 9(2), p. 1.

Crocker, L. & Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart, and Winston.

Cronbach, L. J. & Gleser, G. C. (1957). Psychological tests and personnel decisions. Urbana, IL: University of Illinois Press.

Cronbach, L. J. & Gleser, G. C. (1965). Psychological tests and personnel decisions (2nd ed.). Urbana, IL: University of Illinois Press.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), Educational measurement (2nd ed.) (pp. 499-503). Washington, DC: American Council on Education.

Dodd, B. G. & Kelley, H. P. (1983). SAT-ACT equivalent scores at The University of Texas at Austin, 1980-1983. (Research Bulletin RB-83-9). Austin: University of Texas at Austin, Measurement and Evaluation Center.

Dunn, J. E. (1966). A study of the University of Arkansas mathematics entrance exam as a placement device. The Journal of Experimental Education, 34(3), 62-68.

Duran, N. M. (1985). A theoretical unification of selected aptitude-treatment interaction results. (Doctoral dissertation, The University of Texas at Austin, 1985). Dissertation Abstracts International, 47, 112A.

Ervin, L., Hogrebe, M. C., Dwinell, P. L., & Newman, I. (1984). Comparison of the prediction of academic performance for college developmental students and regularly admitted students. Psychological Reports, 54, 319-327.

Eshenroder, A. L. (1987). Predictors of success in college developmental and technical mathematics courses. (Doctoral dissertation, University of Toledo, 1986). Dissertation Abstracts International, 47, 2217A.

Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. Metron, 1(4), 3-32.

Gagne, R. M. (1970). Conditions of learning (2nd ed.). New York: Holt, Rinehart, & Winston.

Gamache, L. M. & Novick, M. R. (1985). Choice of variables and gender differentiated prediction within selected academic programs. Journal of Educational Measurement, 27(1), 53-70.

Greeno, J. G. (1978). Understanding and procedural knowledge in mathematics instruction. Educational Psychologist, 12(2), 262-283.

Hills, J. R. (1971). Use of measurement in selection and placement. In R. L. Thorndike (Ed.), Educational measurement (2nd ed.) (pp. 600-709). Washington, DC: American Council on Education.

Hunt, R. W. (1987). Predicting relationships between high school mathematics and success in college algebra at the University of Missouri-Columbia. (Doctoral dissertation, University of Missouri-Columbia, 1987). Dissertation Abstracts International, 47, 2268A.

Kelley, H. P. (1979). Defining and applying academic policies about the use of tests for college placement and exemption (Special Report SR-79-12). Austin: The University of Texas at Austin, Measurement and Evaluation Center.

Kirk, R. E. (1982). Experimental design: Procedures for the behaviorial sciences (2nd ed.). Belmont, CA: Wadsworth.

Ku, E. J. & Frisbie, D. A. (1978, March). Methods of evaluating course placement systems. Paper presented at the annual meeting of the National Council on Measurement in Education, Toronto, Canada.

Langston, I. W. & Watkins, T. B. (1980). SAT-ACT Equivalents (Research Memorandum 80-5) . Champaign, IL: University of Illinois, University Office of School and College Relations.

Livingston, S. A. & Zieky, M. J. (1982). Passing scores. Princeton, NJ: Educational Testing Service.

Mathematical Association of America (MAA), (1983, Spring). MAA mini-course on placement testing. PTP Newsletter, 6(2), p. 6. (ERIC Document Reproduction Service No. ED 294 743)

McComb, J. A. (1984). Mathematics placement procedures and psychometric decision theory. (Doctoral dissertation, Michigan State University, 1983) Dissertation Abstracts International, 44, 2742A.

McKillip, W. D. (1966). The effects of high school calculus on student's first semester calculus grades at the University of Virginia. The Mathematics Teacher, 59(5), 470-472.

Morgan, W. P. (1970). Prediction of success in junior college mathematics. The Mathematics Teacher, 63(3), 260-263.

Morris, C. N. & Rolph, J. E. (1981). Introduction to data analysis and statistical inference. Englewood Cliffs, NJ: Prentice-Hall.

Nedelsky, L. (1954). Absolute grading standards for objective tests. Educational and Psychological Measurement, 14(1), 3-19.

Owens, E. M. (1986). Prediction of success in entry level mathematics courses at Tri-County Technical College. (Doctoral dissertation, Clemson University, 1986). Dissertation Abstracts International, 47, 1584A

Romberg, T. A. & Wilson, J. W. (1969). The development of tests (NSLMA report No. 7). Stanford, CA: School Mathematics Study Group.

Rummel, R. J. (1970). Applied factor analysis. Evanston: Northwestern University Press.

SAS (1985). Statistical Analysis System [Computer program] Cary, North Carolina: SAS Institute Inc.

Schallert, D. L. (1982). The significance of knowledge: A synthesis of research related to schema theory. In Wayne Otto and Sandra White (Eds.), Reading Expository Material, (pp. 13-48). New York: Academic Press.

Schoenfeld, A. H. (Ed.) (1987). Cognitive science and mathematics education. Hillsdale, NJ: Lawrence Erlbaum Associates.

Silver, E. A. (1987). Foundations of cognitive theory and research for mathematics problem solving instruction. In A. H. Schoenfeld (Ed.), Cognitive science and mathematics education (pp. 38-60). Hillsdale, NJ: Lawrence Erlbaum Associates.

Skemp, R. R. (1979). Intelligence, learning, and action: A foundation for theory and practice in education. New York: Wiley.

Skemp, R. R. (1987). The psychology of learning mathematics. Hillsdale, NJ: Laurence Erlbaum Associates.

SPSS-X (1986). Statistical package for the social sciences (rels 3.0) [computer program]. Chicago, IL: SPSS Inc.

Tinto, V. (1982). Limits of theory and practice in student attrition. Journal of Higher Education, 53(6), 687-699.

Thorndike, R. L. (1949). Personnel selection: test and measurement techniques. New York: John Wiley & Sons.

Wick, M. E. (1965). A study of factors associated with success in first-year college mathematics. The Mathematics Teacher, 58(7), 642-648.

Willingham, W. W. (1973). The source book for higher education. New York: College Entrance Examination Board.

Willingham, W. W. (1974). College Placement and Exemption. New York: College Entrance Examination Board.

Wilson, J. W. (1971). Evaluation in learning in secondary school mathematics. In B. S. Bloom, J. T. Hastings, & G. F. Madaus (Eds.), Handbook for Formative and Summative Evaluation of Student Learning. (pp. 643- 697). New York: McGraw-Hill.

VITA

Frank Joseph Swehosky [redacted] After completing his work at Lebanon Union High School, Lebanon, Oregon, in 1971, he entered Oregon State University in Corvallis, Oregon. He received his degree of Bachelor of Science from Oregon State University in December, 1975 and also received his commission as a Second Lieutenant in the United States Air Force Reserve. In 1976 he taught mathematics and science at St. John's College, Braybrook, Victoria, Australia. On January 9, 1977 he reported for active duty in the Air Force as a computer programmer/analyst at the 20th Missile Warning Squadron at Eglin AFB, Florida. In the summer of 1978, he received his degree of Master of Science in Science Education from Oregon State University. During May, 1980 to May, 1982 he taught mathematics at the U. S. Air Force Academy Preparatory School. From May, 1982 to July, 1984 he was a mathematics instructor at the U. S. Air Force Academy, Department of Mathematical Sciences. He was subsequently reassigned to the Air Force Operational Test and Evaluation Center, OLDD, at the Lyndon Baines Johnson Space Center, Houston, Texas as a scientific analyst, reporting on the Mission Operations System of the National Space Transportation System (Space Shuttle). He entered The Graduate School of The University of Texas at Austin in September, 1986. He married [redacted]

Publications include:

(1988). Instructional mode consistency. In M. Miller-Gerson (Ed.), Proceedings of the Eighth Annual Microcomputers in Education Conference (pp. 236-240).

(With J. W. Barnes and M. Laguna-Castillo) (1987). Teaching Statistics instruction on a microcomputer local area network. Technology Horizons in Education Journal, 16(2), 80-84.

(With J. S. Gustafson, H. L. Hardwick, L. E. D'Gornaz, D. W. Ehrler, S. Baldwin, G. P. Jenkins, J. Borsi, and P. Brannen) (1986). Air Force operational assessment of the Space Transportation System Mission Operations System (AFOTEC project 0104). Albuquerque, NM: Air Force Operational Test and Evaluation Center.

(With L. E. D'Gornaz, D. W. Ehrler, and S. Baldwin) (1986). Space Transportation System Mission Operations System interim capacity model validation/acceptance test report (AFOTEC project 0104). Albuquerque, NM: Air Force Operational Test and Evaluation Center.

(With B. L. Pasierb) (1984). The Math 130 experiment: A comparison of self-paced and lecture methods of instruction in Pre-calculus (USAFA TR 85-1). Colorado Springs, CO: U. S. Air Force Academy.

In addition, the author has prepared several book reviews that were subsequently published in the following journals: The Mathematics Teacher, The Journal of Computers in Mathematics and Science Teaching, and School Science and Mathematics.

Permanent address: P. O. Box 470
Riddle, Oregon
97469

This dissertation was typed by the author.